PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/57, 9/64, C07K 16/40, C12Q 1/68

A2

(11) International Publication Number: **WO 00/42201**

(43) International Publication Date: 20 July 2000 (20.07.00)

(21) International Application Number: PCT/US00/00641

(22) International Filing Date: 11 January 2000 (11.01.00)

(30) Priority Data:

60/172,247	11 January 1999 (11.01.99)	US
60/132,253	3 May 1999 (03.05.99)	US
60/136,653	27 May 1999 (27.05.99)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	60/172,247 (CIP)
Filed on	11 January 1999 (11.01.99)
US	60/132,253 (CIP)
Filed on	3 May 1999 (03.05.99)
US	60/136,653 (CIP)
Filed on	27 May 1999 (27.05.99)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LAL, Preeti [US/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Park, San Jose, CA 95136 (US).

(74) Agents: HAMLET–COX, Diana et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN PEPTIDASES

(57) Abstract

The invention provides human peptidases (HPEP) and polynucleotides which identify and encode HPEP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of HPEP.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

HUMAN PEPTIDASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human peptidases and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, and metabolic disorders.

BACKGROUND OF THE INVENTION

Peptidases, also called proteases, are enzymes which cleave the peptide bonds forming the backbones of peptides and proteins. Peptidases are required to control the turnover of cellular proteins, which typically have half-lives ranging from hours to a few days. The cleavage of peptide bonds within cells is necessary for the maturation of precursor proteins to their active forms, the removal of signal sequences from targeted proteins, and the degradation of incorrectly folded proteins. Regulated proteolysis and protein degradation by peptidases are essential for normal cell growth, embryonic development, differentiation, wound healing, tissue remodeling, apoptosis, and homeostasis, as well as inflammation and immune response. Peptidases are necessary components of bacterial, parasitic, and viral invasion and replication within a host. Mammalian peptidases have been identified and categorized based on active site structure, mechanism of action, and three-dimensional structure. (See, e.g., Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

The serine proteases (SPs) are a large family of peptidases that include the digestive enzymes trypsin and chymotrypsin; components of the complement and blood-clotting cascades; and enzymes that control the degradation and turnover of macromolecules of the extracellular matrix. SPs are so named because of the presence of a serine residue, usually within a conserved sequence, in the catalytic active site. This catalytic serine forms a triad together with an aspartate and a histidine residue. The main SP sub-families are trypases, which cleave peptide backbones after an arginine or a lysine residue; aspartases, which cleave after aspartate; chymases, which cleave after phenylalanine or leucine; metases, which cleavage after methionine; and serases, which cleave after serine. Pancreatic serine proteases are secreted from the pancreas into the duodenum where they degrade proteins ingested in food. Examples of these proteases include chymotrypsin, trypsin, elastase, and pancreatic kallikrein. Prolylcarboxypeptidase, a lysosomal SP that cleaves peptides such as angiotensin II and III and [des-Arg9] bradykinin, shares sequence homology with members of both the serine carboxypeptidase and prolylendopeptidase families (Tan, F. et al. (1993) J. Biol. Chem. 268:16631-16638). Plasma serine proteases, which include thrombin and C1r, are involved in blood coagulation and immune response. Thrombin converts fibrinogen, a large soluble plasma protein, into

fibrin, a smaller insoluble protein that aggregates to form blood clots. C1r is a component of the complement system, a complex of proteins that perforates the cell membranes of invading microorganisms.

Defects in SPs or their associated regulatory factors are involved in a range of human diseases, including hemorrhagic disorders, thrombophilia, immune disorders, and pancreatic deficiency. For example, mutations in a serine protease cofactor, factor VIII, are the cause of hemophilia. In contrast, excessive expression of the SP prothrombin is one cause of thrombophilia, a genetic predisposition to develop blood clots (Kato, G.J. (1999) Hum. Mutat. 13:87-98). Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by protease cascades. For example, trypsinogen is converted to its active form, trypsin, by enterokinase. Enterokinase, the initiator of intestinal digestion, is an SP found in the intestinal brush border, where it removes an N-terminal fragment from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) Proc. Natl. Acad. Sci. USA 91:7588-7592). In turn, trypsin activates the precursors of the other pancreatic enzymes. Mutations in enterokinase result in severe pancreatic exocrine deficiency (Kato, supra).

The cysteine proteases (CPs) are peptidases involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. CPs have a cysteine as the major catalytic residue in an active site where catalysis proceeds via a thiol ester intermediate and is facilitated by adjacent histidine and aspartic acid residues. Mammalian CPs include lysosomal cathepsins and cytosolic calcium activated proteases (calpains). Cysteine proteases are produced by monocytes, macrophages and other cells of the immune system which migrate to sites of inflammation and, in their protective role, secrete various molecules to repair damaged tissue. Without proper regulation, these cells may overproduce the same molecules and cause tissue destruction in certain disorders. In autoimmune diseases such as rheumatoid arthritis, the secretion of the cysteine protease cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. The cathepsin family of lysosomal proteases includes cysteine proteases (cathepsins B, H, K, L, O2, and S) and aspartyl proteases (cathepsins D and E). Various members of this endosomal peptidase family are differentially expressed. Some, such as cathepsin D, have a ubiquitous tissue distribution while others, such as cathepsin L, are found only in monocytes, macrophages, and other cells of the immune system.

Aspartic proteases (APs) are distinguished from the SPs and CPs by the presence of a pair of aspartic acid residues in the active site, and are most active in the pH 2-3 range, in which one of the aspartate residues is ionized, and the other aspartate is not ionized. APs include penicillopepsin, mammalian pepsin, pepsin A, gastricsin, chymosin, renin, certain fungal peptidases, and members of the cathepsin family of lysosomal proteases such as cathepsins D and E.

Metalloproteases are peptidases which use zinc as an active site component. The zinc atoms of metalloproteases are bound into the enzyme active site by two glutamic acid residues and one histidine residue. Metalloproteases are most notably represented in mammals by the exoproteases carboxypeptidase A and B, and the matrix metalloproteases (MMPs). Carboxypeptidases A and B have similar structures and active sites. Carboxypeptidase A, like chymotrypsin, prefers C-terminal aromatic and aliphatic side chains of hydrophobic nature, whereas carboxypeptidase B is directed toward basic arginine and lysine residues. Another metalloprotease is glycoprotease (GCP), or O-sialoglycoprotein endopeptidase, a peptidase which specifically cleaves O-sialoglycoproteins such as glycophorin A. Placental leucine aminopeptidase (P-LAP) is a metalloprotease which degrades several peptide hormones such as oxytocin and vasopressin, suggesting a role in maintaining homeostasis during pregnancy, and is expressed in several tissues, some of which express two forms of P-LAP mRNAs (Rogi, T. et al. (1996) J. Biol. Chem. 271:56-61).

MMPs are a family of endopeptidases that play an important role in remodeling of the extracellular matrix (ECM). This family includes the collagenases, gelatinases, and stromelysins. MMPs are involved in both normal and pathological tissue remodeling processes including wound healing, inflammation, post-lactational mammary gland involution, and trophoblast invasion during implantation. (See, e.g., Shapiro, S.D. (1998) Curr. Opin. Cell Biol. 10:602-608; Birkedal-Hansen, H. (1995) Curr. Opin. Cell Biol. 7:728-735.) MMPs contribute to the progression of various diseases including arthritis, atherosclerosis, and cancer. MMPs are key players in the irreversible degradation of the ECM seen in rheumatic disease. In cells isolated from inflamed synovia, the mRNAs for stromelysin, cytokines, TIMP-1, cathepsin, gelatinase, and other molecules are preferentially expressed (Keyszer, G.M. (1995) Arthritis Rheum. 38:976-984). A genetic polymorphism which causes diminished expression of stromelysin-1 is associated with enhanced progression of atherosclerosis, a chronic inflammatory process in which plaques are formed in the arterial vessel walls by the accumulation of ECM, smooth muscle cells, and lipid-laden macrophages (Ye, S. et al. (1996) J. Biol. Chem. 271:13055-13060). MMPs play a critical role in tumor invasion and metastasis, helping the tumor to spread by breaking down the surrounding ECM. Overexpression of MMP-3 in mice leads to an increased incidence of breast cancers, while deletions of MMPs suppress tumorigenesis (Sympson, C.J. et al. (1995) Semin. Cancer Biol. 6:159-163; Shapiro, supra). Synthetic MMP inhibitors are currently being tested in clinical trials against breast cancer (Brown, P.D. (1998) Breast Cancer Res. Treat. 52:125-136).

MMPs are regulated in cells by the tissue inhibitors of metalloproteinases (TIMPs). Mutations in TIMP-3 in humans lead to Sorsby's fundus dystrophy, a hereditary degenerative disease of the retina (Weber, B.H. et al. (1994) Nat. Genet. 8:352-356). TIMPs are involved in inhibition of tumor invasion, as overexpression of TIMPs can decrease tumor progression in animal models, and

TIMPs also play a role in regulation of cell growth (Shapiro, supra; Birkedal-Hansen, supra). Overexpresssion of TIMP-3 inhibits tumor invasion in vitro and promotes cell death of different cancer cell types, making it potentially useful for gene therapy of multiple cancer types (Baker, A.H. et al. (1999) Br. J. Cancer 79:1347-1355).

Characteristic sequence motifs in addition to the conserved active site motifs are observed in peptidases. Some SPs contain Kringle domains, triple-looped disulfide cross-linked domains that may function in binding membranes, other proteins or phospholipids, or in the regulation of proteolytic activity. Two plasma serine proteases, plasma kallikrein and coagulation factor XI, have a C-terminal catalytic domain and four tandem N-terminal repeats of about 90 amino acids, including 6 conserved cysteines. Three disulfide bonds linking the first and sixth, second and fifth, and third and fourth cysteines to produce a globular "apple domain."

As an alternative to structure-based classification, peptidases may also be classified by function. Functional classes include the aminopeptidases and signal peptidases. Aminopeptidases catalyze the hydrolysis of amino acid residues from the amino terminus of peptide substrates. Bovine leucine aminopeptidase is a zinc metalloprotease that utilizes the sulfydryl groups from at least three reactive cysteine residues at its active site in the binding of metal ions (Cuypers, H.T. et al. (1982) J. Biol. Chem. 257:7086-7091). Signal peptidases are a specialized class of peptidases that serve in the processing of signal peptides, the amino-terminal sequences which direct a protein from its ribosomal assembly site to a particular cellular or extracellular location. After export, a signal peptidase removes the signal sequence. Signal peptidases exist as multi-subunit complexes in both yeast and mammals.

The ubiquitin-proteasome pathway regulates the proteolysis of cell cycle and growth regulators, including mitotic cyclic kinases; components of signal transduction pathways, including cell surface receptors; transcriptional regulators; oncoproteins; tumor suppressor genes such as p53; viral proteins; and mutated or damaged proteins (Ciechanover, A. (1994) Cell 79:13-21). The system also processes antigens for presentation by the major histocompatability complex class I molecules. Proteins are targeted for degradation by the covalent attachment of multiple molecules of ubiquitin, a small, heat-stable protein, to a lysine residue on the target protein. Attachment of ubiquitin to target proteins is mediated by a member of the ubiquitin ligase family. The ubiquitin-tagged proteins are then recognized and degraded by the proteasome, a large (~2000 kDa), multisubunit complex composed of a central catalytic core containing a variety of peptidases and terminal subunits that serve in substrate recognition and regulation of proteasome activity. During this process, ubiquitin is released from the target proteins and reutilized.

Proteins involved in the ubiquitin-proteasome pathway have been implicated in specific diseases. Certain cell cycle regulators are recognized by multisubunit ubiquitin ligase complexes that

include F-box domain proteins which mediate the recruitment of specific substrates for ubiquitination. Mutations in the ubiquitin ligase enzyme E6-AP are the cause of Angelman's syndrome, a neurological disorder characterized by mental retardation, seizures, and poor coordination and muscle tone. E6-AP is also the target of E6, a viral protein, produced by strains of the human papilloma virus, associated with cervical cancer. E6 modifies the function of E6-AP to accelerate the degradation of the tumor suppressor protein p53 (Ciechanover, A. (1998) EMBO J. 17:7151-7160; Kato, G.J. (1999) Hum. Mutat. 13:87-98). A murine proto-oncogene, Unp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells, and the human homolog of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) Oncogene 10:2179-2183).

Protease inhibitors play a major role in the regulation of the activity and effect of peptidases. For example, the secretory leukocyte protease inhibitor (SLPI) is secreted by epithelial cells and neutrophils, and inhibits leukocyte-secreted serine proteases including elastase and cathepsin G from neutrophils, chymase and trypsin from mast cells, and trypsin and chymotrypsin from pancreatic acinar cells. SLPI and related protease inhibitors are characterized by a four disulfide core structure, or whey acidic protein (WAP) domain. SLPI suppresses the macrophage response to bacterial lipopolysaccharide, which can cause tissue injury, circulatory failure, multiple organ failure, and death. Together with α-1 protease inhibitor, SLPI protects the lungs from emphysema induced by neutrophil elastase. SLPI also possesses antimicrobial activity against fungi, bacteria and HIV (Jin, F.-Y. et al. (1997) Cell 88:417-426; Tomee, J.F. et al. (1998) Thorax 53:114-116).

Cystatins, inhibitors of cysteine proteases, have been associated with a variety of disorders. Low levels of cystatins seem to be correlated with malignant progression of tumors (Calkins, C. et al. (1998) J. Histochem. Cytochem. 46:745-751; Hoppe-Seyler, F. and K.J. Butz (1995) J. Mol. Med. 73:529-538). Increased cysteine protease levels, when accompanied by reductions in inhibitor activity, are correlated with increased malignant properties of tumor cells and the pathology of arthritis and immunological diseases.

The discovery of new human peptidases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, and metabolic disorders.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, human peptidases, referred to collectively as "HPEP" and individually as "HPEP-1," "HPEP-2," "HPEP-3," "HPEP-4," "HPEP-5," "HPEP-6," "HPEP-7," "HPEP-8," "HPEP-9," "HPEP-10," "HPEP-11," "HPEP-12," "HPEP-13," "HPEP-14," "HPEP-15," "HPEP-16," "HPEP-17," and "HPEP-18." In one aspect, the invention provides an

isolated polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-18.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:19-36.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an

amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

The invention further provides an isolated polynucleotide comprising a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide sequence complementary to a), or d) a polynucleotide sequence complementary to b). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide sequence complementary to a), or d) a polynucleotide sequence complementary to b). The method comprises a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 30 contiguous nucleotides. In another alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and a pharmaceutically acceptable excipient. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional HPEP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring amino acid sequence having at least 90% sequence identity

to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional HPEP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional HPEP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:19-36, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

BRIEF DESCRIPTION OF THE FIGURES AND TABLES

Figures 1A, 1B, 1C, 1D, and 1E show the amino acid sequence alignment between HPEP-1 (Incyte Clone ID 155179; SEQ ID NO:1) and human enterokinase (GI 746413; SEQ ID NO:37), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 2A, 2B, and 2C show the amino acid sequence alignment between HPEP-2 (Incyte Clone ID 2415780; SEQ ID NO:2) and Methanococcus jannaschii O-sialoglycoprotein endopeptidase

(GI 2826367; SEQ ID NO:38), produced using the multisequence alignment program of LASERGENE software.

Figures 3A, 3B, and 3C show the amino acid sequence alignment between HPEP-3 (Incyte Clone ID 2879274; SEQ ID NO:3) and human prolylcarboxypeptidase (GI 431321; SEQ ID NO:39), produced using the multisequence alignment program of LASERGENE software.

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding HPEP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of HPEP.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding HPEP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze HPEP, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing

the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"HPEP" refers to the amino acid sequences of substantially purified HPEP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of HPEP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of HPEP either by directly interacting with HPEP or by acting on components of the biological pathway in which HPEP participates.

An "allelic variant" is an alternative form of the gene encoding HPEP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HPEP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as HPEP or a polypeptide with at least one functional characteristic of HPEP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HPEP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HPEP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HPEP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HPEP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide,

polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of HPEP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of HPEP either by directly interacting with HPEP or by acting on components of the biological pathway in which HPEP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind HPEP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the

capability of the natural, recombinant, or synthetic HPEP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding HPEP or fragments of HPEP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser

Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "fragment" is a unique portion of HPEP or the polynucleotide encoding HPEP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:19-36 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:19-36, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:19-36 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:19-36 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:19-36 and the region of SEQ ID NO:19-36 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-18 is encoded by a fragment of SEQ ID NO:19-36. A fragment of SEQ ID NO:1-18 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-18. For example, a fragment of SEQ ID NO:1-18 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-18. The precise length of a fragment of SEQ ID NO:1-18 and the region of SEQ ID NO:1-18 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default

parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a

length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain

DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative

of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of HPEP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HPEP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding HPEP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are

isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned

nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding HPEP, or fragments thereof, or HPEP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence

identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human peptidases (HPEP), the polynucleotides encoding HPEP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, and metabolic disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding HPEP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each HPEP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The Incyte clones in column 5 were used to assemble the consensus nucleotide sequence of each HPEP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

As shown in Figures 1A, 1B, 1C, 1D, and 1E, HPEP-1 has chemical and structural similarity with human enterokinase (GI 746413; SEQ ID NO:37). In particular, HPEP-1 and human enterokinase share 21% identity.

As shown in Figures 2A, 2B, and 2C, HPEP-2 has chemical and structural similarity with Methanococcus jannaschii o-sialoglycoprotein endopeptidase (GI 2826367; SEQ ID NO:38). In particular, HPEP-2 and Methanococcus jannaschii o-sialoglycoprotein endopeptidase share 44% identity.

As shown in Figures 3A, 3B, and 3C, HPEP-3 has chemical and structural similarity with human prolylcarboxypeptidase (GI 431321; SEQ ID NO:39). In particular, HPEP-3 and human prolylcarboxypeptidase share 33% identity.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding HPEP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:19-36

and to distinguish between SEQ ID NO:19-36 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express HPEP as a fraction of total tissues expressing HPEP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing HPEP as a fraction of total tissues expressing HPEP. Of particular note is the expression of SEQ ID NO:28 in tissues associated with inflammation and the immune response. Column 5 lists the vectors used to subclone each cDNA library.

Northern analysis shows the expression of SEQ ID NO:19 in various libraries, at least 66% of which are associated with cell proliferation and at least 31% of which are associated with inflammation and immune response. Of particular note is the expression of HPEP-1 in gastrointestinal tissues (33%), reproductive tissues (28%), and hematopoietic/immune tissues (28%).

Northern analysis shows the expression of SEQ ID NO:20 in various libraries, at least 59% of which are associated with cell proliferation and at least 43% of which are associated with inflammation and immune response. Of particular note is the expression of HPEP-2 in reproductive tissues (21%), hematopoietic/immune tissues (20%), and nervous tissues (19%).

Northern analysis shows the expression of SEQ ID NO:21 in various libraries, at least 61% of which are associated with cell proliferation and at least 34% of which are associated with inflammation and immune response. Of particular note is the expression of HPEP-3 in reproductive tissues (30%), nervous tissues (18%), and gastrointestinal tissues (12%).

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding HPEP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:30 maps to chromosome 17 within the interval from 75.70 to 83.90 centiMorgans. This interval also contains a gene associated with hepatic leukemia and estrogen response. SEQ ID NO:32 maps to chromosome 7 within the interval from 78.90 to 79.60 centiMorgans.

The invention also encompasses HPEP variants. A preferred HPEP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the HPEP amino acid sequence, and which contains at least one functional or structural characteristic of HPEP.

The invention also encompasses polynucleotides which encode HPEP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36, which encodes HPEP.

The invention also encompasses a variant of a polynucleotide sequence encoding HPEP. In

particular, such a variant polynucleotide sequence will have at least about 85%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HPEP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36 which has at least about 85%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:19-36. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HPEP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HPEP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HPEP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode HPEP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring HPEP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HPEP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HPEP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HPEP and HPEP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HPEP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:19-36 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol.

152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding HPEP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in

length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HPEP may be cloned in recombinant DNA molecules that direct expression of HPEP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express HPEP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HPEP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of HPEP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then

subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding HPEP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, HPEP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of HPEP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active HPEP, the nucleotide sequences encoding HPEP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HPEP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HPEP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HPEP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be

provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HPEP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HPEP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding HPEP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding HPEP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding HPEP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of HPEP are needed, e.g. for the production of antibodies, vectors which direct high level expression of HPEP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of HPEP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994)

Bio/Technology 12:181-184.)

Plant systems may also be used for expression of HPEP. Transcription of sequences encoding HPEP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HPEP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses HPEP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of HPEP in cell lines is preferred. For example, sequences encoding HPEP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *aprˉ* cells, respectively. (See, e.g., Wigler, M. et

al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HPEP is inserted within a marker gene sequence, transformed cells containing sequences encoding HPEP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HPEP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding HPEP and that express HPEP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of HPEP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HPEP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and

may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HPEP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HPEP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HPEP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HPEP may be designed to contain signal sequences which direct secretion of HPEP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HPEP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric HPEP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HPEP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate

fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HPEP encoding sequence and the heterologous protein sequence, so that HPEP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled HPEP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

Fragments of HPEP may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of HPEP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of HPEP and human peptidases. In addition, the expression of HPEP is closely associated with cancer and cell proliferation, inflammation and immune response, reproductive tissues, hematopoietic/immune tissues, gastrointestinal tissues, and nervous tissues. Therefore, HPEP appears to play a role in cell proliferative, autoimmune/inflammatory, and metabolic disorders. In the treatment of disorders associated with increased HPEP expression or activity, it is desirable to decrease the expression or activity of HPEP. In the treatment of disorders associated with decreased HPEP expression or activity, it is desirable to increase the expression or activity of HPEP.

Therefore, in one embodiment, HPEP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HPEP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone,

bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a metabolic disorder such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, and pseudovitamin D-deficiency rickets.

In another embodiment, a vector capable of expressing HPEP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HPEP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified HPEP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HPEP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HPEP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HPEP including, but not limited to, those listed above.

In a further embodiment, an antagonist of HPEP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HPEP. Examples of such

disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, and metabolic disorders described above. In one aspect, an antibody which specifically binds HPEP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express HPEP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HPEP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HPEP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HPEP may be produced using methods which are generally known in the art. In particular, purified HPEP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HPEP. Antibodies to HPEP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HPEP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HPEP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of HPEP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to HPEP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HPEP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for HPEP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HPEP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HPEP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for HPEP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of HPEP-antibody complex

divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple HPEP epitopes, represents the average affinity, or avidity, of the antibodies for HPEP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular HPEP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the HPEP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of HPEP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of HPEP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding HPEP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding HPEP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HPEP. Thus, complementary molecules or fragments may be used to modulate HPEP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HPEP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding HPEP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding HPEP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding HPEP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in

the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding HPEP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HPEP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HPEP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HPEP, antibodies to HPEP, and mimetics, agonists, antagonists, or inhibitors of HPEP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration.

Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner

that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HPEP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HPEP or fragments thereof, antibodies of HPEP, and agonists, antagonists or inhibitors of HPEP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the

subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind HPEP may be used for the diagnosis of disorders characterized by expression of HPEP, or in assays to monitor patients being treated with HPEP or agonists, antagonists, or inhibitors of HPEP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HPEP include methods which utilize the antibody and a label to detect HPEP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HPEP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HPEP expression. Normal or standard values for HPEP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to HPEP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of HPEP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HPEP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of HPEP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HPEP, and to monitor regulation of HPEP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HPEP or closely related molecules may be used to

identify nucleic acid sequences which encode HPEP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding HPEP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the HPEP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:19-36 or from genomic sequences including promoters, enhancers, and introns of the HPEP gene.

Means for producing specific hybridization probes for DNAs encoding HPEP include the cloning of polynucleotide sequences encoding HPEP or HPEP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HPEP may be used for the diagnosis of disorders associated with expression of HPEP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic

anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a metabolic disorder such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, and pseudovitamin D-deficiency rickets. The polynucleotide sequences encoding HPEP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered HPEP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HPEP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HPEP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HPEP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HPEP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HPEP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the

patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HPEP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HPEP, or a fragment of a polynucleotide complementary to the polynucleotide encoding HPEP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of HPEP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding HPEP may be used

to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding HPEP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HPEP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HPEP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with HPEP, or fragments thereof,

and washed. Bound HPEP is then detected by methods well known in the art. Purified HPEP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HPEP specifically compete with a test compound for binding HPEP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HPEP.

In additional embodiments, the nucleotide sequences which encode HPEP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No. PF-0651 P, filed January 11, 1999], U.S. Ser. No. 60/132,253, and U.S. Ser. No. 60/136,653, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA

libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides

were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and

amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:19-36. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding HPEP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of HPEP Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:30-36 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:30-36 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available

from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:30 and SEQ ID NO:32 are described in The Invention as ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VI. Extension of HPEP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:19-36 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times;

Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:19-36 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:19-36 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide

fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

IX. Complementary Polynucleotides

Sequences complementary to the HPEP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HPEP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of HPEP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HPEP-encoding transcript.

X. Expression of HPEP

Expression and purification of HPEP is achieved using bacterial or virus-based expression systems. For expression of HPEP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express HPEP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of HPEP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding HPEP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, HPEP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from HPEP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate

resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified HPEP obtained by these methods can be used directly in the following activity assay.

XI. Demonstration of HPEP Activity

Peptidase activity of HPEP is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric or fluorometric absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (Carboxypeptidase A and B, procollagen C-proteinase). Chromogens commonly used are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are performed at room temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette and monitored by measurement of the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate. The change in absorbance is proportional to the peptidase activity of HPEP in the assay.

Alternatively, regulation of peptidase activity (agonism or antagonism) by HPEP is measured using an appropriate protease assay as described above in the presence or absence of HPEP as an agonist or inhibitor of this activity. Protease activity is measured in the absence of HPEP (control activity) and in the presence of varying amounts of HPEP. The change in protease activity compared to the control is proportional to the amount of HPEP in the assay and is a measure of the protease regulatory activity of HPEP.

Alternatively, ubiquitin activity of HPEP is demonstrated by its ability to form a covalent thiolester bond with ubiquitin-activating enzyme (E1). This activity can be detected and quantified using a "covalent affinity" chromatography procedure (Ciechanover, A. et al. (1982) J. Biol. Chem. 257:2537-2542). E1 is first conjugated to SEPHAROSE resin, an inert resin, using methods well known by those skilled in the art. HPEP, produced by recombinant methods or purified biochemically, is present in a solution containing ATP and magnesium ions. This solution is exposed to the E1-Sepharose conjugate in a column chromatography format. E1-Sepharose is washed with a solution containing a high concentration of salt, such as sodium chloride. This treatment is effective in removing virtually all proteins that are not covalently bound to E1-Sepharose. HPEP covalently bound to E1-Sepharose is eluted with a thiol compound such as dithiothreitol. The presence of HPEP in the eluent is detected by SDS-polyacrylamide gel electrophoresis and gel staining. Immunological methods such as western blot which utilize specific antibody directed against HPEP are used to quantify the amount of HPEP in the eluent. The amount of HPEP that binds to E1-Sepharose is

proportional to the ubiquitin activity of HPEP.

XII. Functional Assays

HPEP function is assessed by expressing the sequences encoding HPEP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of HPEP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding HPEP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding HPEP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of HPEP Specific Antibodies

HPEP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the HPEP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-HPEP activity by, for example, binding the peptide or HPEP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring HPEP Using Specific Antibodies

Naturally occurring or recombinant HPEP is substantially purified by immunoaffinity chromatography using antibodies specific for HPEP. An immunoaffinity column is constructed by covalently coupling anti-HPEP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HPEP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HPEP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HPEP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HPEP is collected.

XV. Identification of Molecules Which Interact with HPEP

HPEP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HPEP, washed, and any wells with labeled HPEP complex are assayed. Data obtained using different concentrations of HPEP are used to calculate values for the number, affinity, and association of HPEP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are

obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	19	155179	THP1PLB02	155179H1 (THP1PLB02), 155179X307D2 (THP1PLB02), 1214111X24 (BRSTTUT01), 961990X21 (BRSTTUT03), 961990X17 (BRSTTUT03), 478199X15 (MMLR2DT01), 692774X19 (LUNGTUT02), 034109F1 (THP1NOB01), 2754059H1 (THP1AZS08)
2	20	2415780	HNT3AZT01	2415780H1 (HNT3AZT01), 1443076F6 (THYRNOT03),1753156F6 (LIVRTUT01), 989015H1 (LVENNOT03), 1922409R6 (BRSTTUT01)
3	21	2879274	UTRSTUT05	2879274H1 (UTRSTUT05), 3537571H1 (SEMVNOT04), 2879274H1 (UTRSTUT05), 2767241H1 (COLANOT02), 1479540F1 (CORPNOT02), 1650591F6 (PROSTUT09), 1650591T6 (PROSTUT09), 1264516R1 (SYNORAT05), 1438281F1 (PANCNOT08)
4	22	358050	PROSNOT01	041451R6 (TBLYNOT01), 358050H1 (PROSNOT01), 1288739F6 (BRAINOT11), 1338092F6 (COLNNOT13), 1338092T6 (COLNNOT13), 1817810F6 (PROSNOT20), 3049061H1 (LUNGNOT25), 3217540H1 (TESTNOT07), 3224582H2 (UTRSNON03)
5	23	700745	SYNORAT03	700745H1 (SYNORAT03), 700745R6 (SYNORAT03)
6	24	2026480	KERANOT02	1288279F6 (BRAINOT11), 1798769H1 (COLNNOT27), 1984648T6 (LUNGAST01), 2026480H1 (KERANOT02), 3577373F6 (BRONNOT01), 4049569T6 (SINTNOT18), SAYA00492F1
7	25	2132401	OVARNOT03	014071R6 (THP1PLB01), 2669596F6 (ESOGTUT02), 4511344H1 (EPIMNOT01), SAJA01969F1, SAJA00384R1, SAJA00561F1
8	26	2568875	HIPOAZT01	826204R1 (PROSNOT06), 826204X144F1 (PROSNOT06), 826204X48 (PROSNOT06), 826204X52 (PROSNOT06), 2568875H1 (HIPOAZT01)
9	27	3408908	PROSTUS08	3408908F6 (PROSTUS08), 3408908H1 (PROSTUS08), SBWA03204V1
10	28	3772696	BRSTNOT25	3772696F6 (BRSTNOT25), 3772696H1 (BRSTNOT25), 3772696T6 (BRSTNOT25), SXBA00825V1, SXBA00411V1

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
11	29	5388674	BRAINOT19	2260285X320D4 (UTRSNOT02), 2287941X301D1 (BRAINON01), 2289534R6 (BRAINON01), 3521165H1 (LUNGNON03), 4315221H1 (BRAFNOT01), 5082079H1 (LNODNOT11), 5388674H1 (BRAINOT19)
12	30	1873102	LEUKNOT02	092669F1 (HYPONOB01), 225519F1 (PANCNOT01), 225519R1 (PANCNOT01), 390991H1 (TMLR2DT01), 1737263F6 (COLNNOT22), 1737263T6 (COLNNOT22), 1873102H1 (LEUKNOT02), 1932133F6 (COLNNOT16), 3590995H1 (293TF5T01), 3712151H1 (PENCNOT09), 4285941H1 (LIVRDIR01), 4339405H1 (BRAUNOT02)
13	31	1920734	BRSTTUT01	991651H1 (COLNNOT11), 1920734H1 (BRSTTUT01), 1920734R6 (BRSTTUT01), 1920734T6 (BRSTTUT01), 2739282F6 (OVARNOT09), 3765480H1 (BRSTNOT24)
14	32	2396858	THP1AZT01	1439237F1 (PANCNOT08), 1722122F6 (BLADNOT06), 1908978F6 (CONNTUT01), 2396858H1 (THP1AZT01), 2396858X301V1 (THP1AZT01), 2396858X305D1 (THP1AZT01), 2461972F6 (THYRNOT08)
15	33	2634725	COLNTUT15	1875442H1 (LEUKNOT02), 2634725F6 (COLNTUT15), 2634725H1 (COLNTUT15), 2920995T6 (SININOT04), 4875374H1 (COLDNOT01)
16	34	2643110	LUNGTUT08	881275H1 (THYRNOT02), 1273883X302D2 (TESTTUT02), 1273883X304D2 (TESTTUT02), 1918031R6 (PROSNOT06), 2171263F6 (ENDCNOT03), 2453207F6 (ENDANOT01), 2453207T6 (ENDANOT01), 2643110H1 (LUNGTUT08), 2753878H1 (THP1AZS08)
17	35	2701396	OVARTUT10	2701396H1 (OVARTUT10), 2867440T6 (KIDNNOT20), SBLA01199F1, SBLA03620F1, SBLA02714F1
18	36	3134404	SMCCNOT01	3134404H1 (SMCCNOT01), 4161423F6 (BRSTNOT32), 4384476H1 (BRAVUTT02), SCAA06693V1

Table 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
1	762	S174 S190 T225 S341 S417 T526 T681 S717 S732 S94 T166 T227 S329 T333 T365 T368 T402 T436 T478 S589 T692 Y342	N16 N209 N392 N679	Cell attachment sequence: R156-D158 Serine protease trypsin family active sites: C548-C564, V559-C564, V654-T660, D706-S717, C708-S718, W733-I756 Trypsin motif: V522-I756 Chymotrypsin serine protease family: G549-C564, D614-A628, V705-S717 Low-density lipoprotein receptor: D358-C395, G369-E390, C371-E390, A397-C432, K406-E427, P433-K468, G442-E463, V472-C511 Kringle motif: C548-Y565, D705-S718 Developmental CUB domain: C121-F238, C247-Y351	Epithin (membrane bound serine protease) [Mus musculus] g4104970 Enterokinase [Homo sapiens] g746413 (21% identity)	MOTIFS BLOCKS PRINTS PFAM BLAST

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
2	335	T67 S209 T225 S232 T45 T196 T245 T321 T332 Y195		Glycoprotease family: V4-V18, T38-L66, D70-I114, R94-I114, T127-I139, Y146-I170, C252-E271 Glycoprotease motif: P2-R308 O-sialoglycoprotein endopeptidase: L5-V18, I74-I114, T127-I139, F149-I170, A257-N266	O-sialoglycoprotease [Rattus norvegicus] g5360708 O-sialoglycoprotein endopeptidase [Methanococcus jannaschii] g2826367 (44% identity)	MOTIFS BLOCKS PRINTS PFAM BLAST
3	327	S152 S166 T175 S285 S292 S48 T73	N150 N191 N198 N263	Signal peptide: M1-A26	Prolylcarboxy-peptidase [Homo sapiens] g431321 (33% identity)	MOTIFS SPScan BLAST
4	471	S44 S468 S26 S47 S64 T82 S117 T244 T280 S445 S40 T69 S145 T307 T405 Y106 Y223		F-box domain: P10-H56 Signal peptide: M1-L33	F-box protein sequence (GeneSeq Y02253)	MOTIFS PFAM SPScan BLAST
5	60	S15		Signal peptide: M1-G20 Tissue inhibitor of metalloproteinases signature: G17-C46	TIMP-3 (Tissue inhibitor of metalloproteinases-3) [Homo sapiens] g1215682	BLAST MOTIFS SPSCAN BLOCKS HMM ProfileScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
6	399	S24 S45 S30 S163 T211 S244 T321			PINT domain protein (Proteasomal subunit) [Plasmodium falciparum] g3845132	BLAST MOTIFS
7	106	S6 S20 T32 T102 S63 S69 T74		Ubiquitin signature: V39-V90		MOTIFS BLOCKS
8	267	S2 T72 T89 S211 S236 S12 S111	N260	Trypsin serine protease active site: L51-C56 Trypsin serine protease signature: T15-V235 Signal peptide: M1-Q61	Prostasin (serine protease) [Homo sapiens] g1143194	BLAST MOTIFS PFAM BLOCKS PRINTS SPSCAN ProfileScan
9	123	S44 S80 T58 S75 S103		Signal peptide: M1-G24 WAP domain: K30-P72, K77-P120	Secretory leukocyte protease inhibitor [Mus musculus] g1763263	BLAST MOTIFS SPScan PFAM PRINTS HMM ProfileScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
10	513	T24 S57 T193 S249 T311 T75 T88 T112 T290 T384 T385 S422	N55 N110 N200 N452 N470 N508	Signal peptide: M1-P19 Matrixin signature: Y36-T202 Neutral zinc metalloproteases Zn-binding region: V213-L222 Hemopexin domain: F285-C465	Matrix metalloproteinase [Gallus gallus] g3511149	BLAST MOTIFS SPScan PFAM BLOCKS PRINTS HMM ProfileScan
11	326	S191 T140 S158 S196 S269 S296 T26 T41 S104 S214	N34	Neutral zinc metalloproteases Zn-binding region: T217-G227		MOTIFS BLOCKS ProfileScan
12	823	T6 S350 S168 T277 S353 S381 S398 T407 S415 S479 S524 S531 S566 T641 T97 T146 T194 T271 T277 T331 S394 T435 T658 T727 T753 T806	N111 N213 N329 N421 N596	Ubiquitin carboxyl-terminal hydrolases family 2: G197-L214, Y295-L304, V355-C369, L741-A765, Y742-Y760, K790-N811 Ubiquitin carboxyl-terminal hydrolase family: G197-L214, Y742-V801	Ubiquitin-specific protease UBP41 [Mus musculus] g3386552	MOTIFS BLAST PFAM BLIMPS
13	404	S58 S68 T107 S164 T177 T208 S284 T14 S68 T341	N339		Similar to zinc metalloprotease [C. elegans] g2804437	MOTIFS BLAST

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
14	703	S20 S68 T120 T135 S331 T383 S562 S606 S607 S631 S674 S698 T31 S95 S115 S173 S355 S490 S562 S650	N318 N434 N445 N670	E1 ubiquitin activating enzyme: K352-H442	E1-like protein (ubiquitin activating enzyme) [Pichia pastoris] g4262402	MOTIFS BLAST BLIMPS
15	145	T36 S100 S115 T47	N34	Protease serine hydrolase precursor signal zymogen glycoprotein multigene family: L16-Q64, G87-K140 Trypsin: L25-Q64, S84-N142	Matriptase (serine protease) [Homo sapiens] g5359675, g6002714 Epithin (membrane bound serine protease) [Mus musculus] g4104970	MOTIFS BLAST BLIMPS
16	518	S74 T252 S151 T169 T245 S312 S361 T419 S462 S502 S16 S70 S98 S133 T301 S331 S428 T516 Y334	N234	Dipeptidyl peptidase IV: H255-L305, E326-Q352, E379-P411	Dipeptidyl peptidase IV [Stenotrophomonas maltophilia] g1753197	MOTIFS BLAST BLIMPS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
17	476	S25 T183 S203 S324 S358 T398 S441 S457 T471 S472 S25 S345 T398 S402 T420 Y34 Y447	N38 N175 N314 N360 N455	Eukaryotic thiol proteases active site: F431-I450 Cysteine protease: C240-W467 Eukaryotic thiol protease active site: K237-F246, R281-I289, T410-G419, F431-Y447	Similar to cysteine protease [C. elegans] g3876422 Cathepsin B [Triticum aestivum] g21693	MOTIFS BLAST BLIMPS PFAM BLIMPS
18	229	S96 S219 S77 S78	N11 N105 N125	Signal peptide: M1-C25 Zn metalloprotease: S63-L210	Zinc metalloprotease ADAMTS7 [Homo sapiens] g5923788	MOTIFS BLAST HMM BLIMPS

Table 3

Nucleotide SEQ ID NO:	Useful Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
19	322-366	Gastrointestinal (0.33) Reproductive (0.28) Hematopoietic/Immune (0.28)	Cell Proliferation (0.66) Inflammation and Immune Response (0.31)	
20	499-543	Reproductive (0.21) Hematopoietic/Immune (0.20) Nervous (0.19)	Cell Proliferation (0.59) Inflammation and Immune Response (0.43)	
21	1082-1126	Reproductive (0.30) Nervous (0.18) Gastrointestinal (0.12)	Cell Proliferation (0.61) Inflammation and Immune Response (0.34)	
22	305-478 1847-1891	Reproductive (0.360) Nervous (0.220) Cardiovascular (0.100)	Cell Proliferation (0.560) Inflammation and Immune Response (0.200)	PBLUESCRIPT
23	146-190	Reproductive (0.500) Developmental (0.250) Musculoskeletal (0.250)	Cell Proliferation (0.250) Inflammation and Immune Response (0.250)	PSPORT1
24	433-477	Reproductive (0.250) Gastrointestinal (0.155) Hematopoietic/Immune (0.155) Nervous (0.155)	Cell Proliferation (0.667) Inflammation and Immune Response (0.274)	PSPORT1
25	56-100 440-484	Gastrointestinal (0.207) Reproductive (0.207) Cardiovascular (0.103) Hematopoietic/Immune (0.103) Musculoskeletal (0.103) Nervous (0.103)	Cell Proliferation (0.589) Inflammation and Immune Response (0.448)	PSPORT1

Table 3 (cont.)

Nucleotide SEQ ID NO:	Useful Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
26	704-748 1001-1045	Reproductive (0.500) Cardiovascular (0.125) Gastrointestinal (0.125) Hematopoietic/Immune (0.125) Nervous (0.125)	Cell Proliferation (0.375) Inflammation and Immune Response (0.250)	PSPORT1
27	189-233 327-371	Reproductive (0.667) Dermatologic (0.333)	Cell Proliferation (0.667) Inflammation and Immune Response (0.333)	PT7T3
28	168-212 1227-1271	Reproductive (1.000)	Inflammation and Immune Response (1.000)	pINCY
29	226-270	Reproductive (0.258) Nervous (0.194) Hematopoietic/Immune (0.172)	Cell Proliferation (0.591) Inflammation and Immune Response (0.376)	pINCY
30	649-693	Reproductive (0.235) Hematopoietic/Immune (0.163) Nervous (0.153)	Cancer (0.418) Inflammation (0.276) Cell Proliferation (0.163)	pINCY
31	379-423	Reproductive (0.348) Nervous (0.217) Cardiovascular (0.174)	Cancer (0.435) Inflammation (0.130) Cell Proliferation (0.087) Trauma (0.087)	PSPORT1
32	704-748	Reproductive (0.262) Hematopoietic/Immune (0.167) Nervous (0.143)	Cancer (0.500) Inflammation (0.262) Cell Proliferation (0.214)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Useful Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
33	433-477	Gastrointestinal (0.365) Reproductive (0.288) Hematopoietic/Immune (0.115)	Cancer (0.538) Inflammation (0.250) Trauma (0.115)	pINCY
34	1398-1442	Reproductive (0.200) Cardiovascular (0.150) Gastrointestinal (0.150) Nervous (0.150)	Cancer (0.350) Cell Proliferation (0.300) Inflammation (0.150)	pINCY
35	755-801	Urologic (0.500) Gastrointestinal (0.167) Nervous (0.167) Reproductive (0.167)	Cancer (0.667) Trauma (0.333)	pINCY
36	447-491	Reproductive (0.375) Cardiovascular (0.250) Developmental (0.125) Nervous (0.125) Urologic (0.125)	Cell Proliferation (0.500) Other (0.250) Inflammation (0.125) Trauma (0.125)	pINCY

Table 4

SEQ ID NO:	Library	Library Comments
19	THP1PLB02	Library was constructed by reamplification of a human promonocyte line library, which was made using RNA isolated from THP-1 cells cultured for 48 hours with 100 ng/ml phorbol ester (PMA), followed by a 4-hour culture in media containing 1 ug/ml LPS. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old male with acute monocytic leukemia.
20	HNT3AZT01	Library was constructed using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated for three days with 0.35 micromolar 5-aza-2'-deoxycytidine (AZ).
21	UTRSTUT05	Library was constructed using RNA isolated from uterine tumor tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated uterine leiomyoma. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Patient history included a ventral hernia and a benign ovarian neoplasm.
22	PROSNOT01	Library was constructed using RNA isolated from the prostate tissue of a 78-year-old Caucasian male, who died from leukemia. Patient history included skin cancer, emphysema, and asthma. Previous surgeries included a cholecystectomy.
23	SYNORAT03	Library was constructed using RNA isolated from the wrist synovial membrane tissue of a 56-year-old female with rheumatoid arthritis.
24	KERANOT02	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is a human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
25	OVARNOT03	Library was constructed using RNA isolated from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comments
26	HIPOAZT01	Library was constructed from RNA isolated from diseased hippocampus tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
27	PROSTUS08	Library was constructed using 2.36 million clones from a prostate tumor library and was subjected to one round of subtractive hybridization with 448,000 clones from a control prostate library. The starting library for subtraction was constructed using RNA isolated from a prostate tumor removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3). Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., Nucleic Acids Res. (1991) 19:1954 and Bonaldo et al., Genome Research (1996) 6:791.
28	BRSTNOT25	Library was constructed using RNA isolated from breast tissue removed from a 35-year-old Caucasian female during a bilateral reduction mammoplasty. Family history included uterine cancer, hyperlipidemia, benign hypertension, acute myocardial infarction, cerebrovascular disease, atherosclerotic coronary artery disease, and type II diabetes.
29	BRAINOT19	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology indicated a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. This tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. The left temporal lobe, including the mesial temporal structures, showed focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. GFAP was positive for astrocytes. The patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, and depressive disorder. Family history included brain cancer.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comments
30	LEUKNOT02	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).
31	BRSTTUT01	Library was constructed using RNA isolated from breast tumor tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated invasive grade 4 mammary adenocarcinoma of mixed lobular and ductal type, extensively involving the left breast. The tumor was identified in the deep dermis near the lactiferous ducts with extracapsular extension. Proliferative fibrocysytic changes were characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Patient history included atrial tachycardia, blood in the stool, and a benign breast neoplasm. Family history included benign hypertension, atherosclerotic coronary artery disease, cerebrovascular disease, and depressive disorder.
32	THP1AZT01	Library was constructed using polyA RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from blood of a 1-year-old Caucasian male with acute monocytic leukemia.
33	COLNTUT15	Library was constructed using RNA isolated from colon tumor tissue obtained from a 64-year-old Caucasian female during a right hemicolectomy with ileostomy and bilateral salpingo-oophorectomy (removal of the fallopian tubes and ovaries). Pathology indicated an invasive grade 3 adenocarcinoma. Patient history included hypothyroidism, depression, and anemia. Family history included colon cancer and uterine cancer.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comments
34	LUNGTUT08	Library was constructed using RNA isolated from lung tumor tissue removed from a 63-year-old Caucasian male during a right upper lobectomy with fiberoptic bronchoscopy. Pathology indicated a grade 3 adenocarcinoma. Patient history included atherosclerotic coronary artery disease, an acute myocardial infarction, rectal cancer, an asymtomatic abdominal aortic aneurysm, and cardiac dysrhythmia. Family history included congestive heart failure, stomach cancer, and lung cancer, type II diabetes, atherosclerotic coronary artery disease, and an acute myocardial infarction.
35	OVARTUT10	Library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 58-year-old Caucasian female during a total abdominal hysterectomy, removal of a solitary ovary, and repair of inguinal hernia. Pathology indicated a metastatic grade 3 adenocarcinoma of colonic origin, forming a partially cystic and necrotic tumor mass in the left ovary, and an adenocarcinoma of colonic origin, forming a nodule in the left mesovarium. A single intramural leiomyoma was identified in the myometrium. The cervix showed mild chronic cystic cervicitis. Patient history included benign hypertension, follicular cyst of the ovary, colon cancer, benign colon neoplasm, and osteoarthritis. Family history included emphysema, myocardial infarction, atherosclerotic coronary artery disease, benign hypertension, and hyperlipidemia.
36	SMCCNOT01	Library was constructed using RNA isolated from smooth muscle cells removed from the coronary artery of a 3-year-old Caucasian male.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-18,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

2. An isolated polypeptide of claim 1, having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide of claim 3, having a sequence selected from the group consisting of SEQ ID NO:19-36.

5. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

6. A cell transformed with a recombinant polynucleotide of claim 5.

7. A transgenic organism comprising a polynucleotide of claim 5.

8. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

9. An isolated antibody which specifically binds to a polypeptide of claim 1.

10. An isolated polynucleotide comprising:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,

b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,

c) a polynucleotide sequence complementary to a), or

d) a polynucleotide sequence complementary to b).

11. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 10.

12. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 10, the method comprising:

a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

13. A method of claim 12, wherein the probe comprises at least 30 contiguous nucleotides.

14. A method of claim 12, wherein the probe comprises at least 60 contiguous nucleotides.

15. A pharmaceutical composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

16. A method of treating a disease or condition associated with decreased expression of functional HPEP, comprising administering to a patient in need of such treatment the pharmaceutical composition of claim 15.

17. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

18. A pharmaceutical composition comprising an agonist compound identified by a method of claim 17 and a pharmaceutically acceptable excipient.

19. A method of treating a disease or condition associated with decreased expression of functional HPEP, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 18.

20. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

21. A pharmaceutical composition comprising an antagonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with overexpression of functional HPEP, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 21.

23. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 4, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

```
1    M------RITNENFVDAYE--------------- 155179
1    MGSKRGISSRHHSLSSYEIMFAALFAILVV g746413

14   ----------------------------NSNST 155179
31   LCAGLIAVSCLTIKESQRGAALGQSHEARA g746413

19   EF------VSLASKVKDALKLLYSGVPF--- 155179
61   TFKITSGVTYNPNLQDKLSVDFKVLAFDLQ g746413

41   ------------LGPCHKESAVTAFSEGSVI 155179
91   QMIDEIFLSSNLKNEYKNSRVLQFENGSII g746413

60   AYY-------W-SEFSIPQHLVEEAERVMAE 155179
121  VVFDLFFAQWVSDQNVKEELIQGLEANKSS g746413

83   ERVVMLPP--RARSLKSFVVTSVVAFPTDS 155179
151  QLVTFHIDLNSVDILDKLTTTSHLATPGNV g746413

111  KTVQRTQDNSCSFGLHARGVELM------- 155179
181  SIECLPGSSPCTDALTCIKADLFCDGEVNC g746413

134  ----------------------RF----TTPGFPD 155179
211  PDGSDEDNKMCATVCDGRFLLTGSSGSFQA g746413
```

FIGURE 1A

```
143  S P Y P A H A R - - - - C Q W A L R G D A D S V L S L T F R  155179
241  T H Y P K P S E T S V V C Q W I I R V N Q G L S I K L S F D  g746413

169  S F D - - - - - - - L A S C D E R G S D L V T V Y N T L S P M  155179
271  D F N T Y Y T D I L D I Y E G V G S S K I - L R A S I W E T  g746413

193  E P H A L V Q L C G T Y P P S Y - - - - - - - - - N L T F H S  155179
300  N P G T I R I F S N Q V T A T F L I E S D E S D Y V G F N A  g746413

215  S Q N V L L I T L I T N T E R R H P G F E A T F - F Q L P R  155179
330  T Y T A F N S S E L N N Y E K I N C N F E D G F C F W V Q D  g746413

244  M S S C G G R L R K A Q G T F N S P Y Y P G H Y P P N I D -  155179
360  L N D D N E W E R I Q G S T F - S P F T G P N F D H T F G N  g746413

273  - C T W N I E V P N N Q H V K V - R F K F F Y L - L E P G V  155179
389  A S G F Y I S T P T G P G G R Q E R V G L L S L P L D P T L  g746413

300  P A G T C P K D Y V E I N G E K Y C G E R S Q F V V T S N S  155179
419  E P A - C L S F W Y H M Y G E N - - V H K L S I N I S N D Q  g746413

330  N K I T V R F H S D Q S Y T D T - - - - - - - - - - - - - G F  155179
446  N M E K T V F Q K E G N Y G D N W N Y G Q V T L N E T V K F  g746413
```

FIGURE 1B

```
348  L A E Y L S Y D S - - - S D P C P G Q F T C R T G R C I R K  155179
476  K V A F N A F K N K I L S D I A L D D I S L T Y G I C - N G  g746413

375  E L R C D G W A D C T D H S D E L N C S C D A G H Q F - T C  155179
505  S L Y P E P - T L V P T P P P E L P T D C G G P F E L W E P  g746413

404  K N K F C K P L F - - - - - - - - - - - - W V C - - - - - -  155179
534  N T T F S S T N F P N S Y P N L A F C V W I L N A Q K G K N  g746413

416  - - - - - - - - - - - D S V N D C G D N S D E Q G C - - - - -  155179
564  I Q L H F Q E F D L E N I N D V V E I R D G E E A D S L L L  g746413
```

```
431  - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  155179
594  A V Y T G P G P V K D V F S T T N R M T V L L I T N D V L A  g746413

431  - - - - - - - - - - - - - - - - - - - - S C P A Q T F R C S N G  155179
624  R G G F K A N F T T G Y H L G I P E P C K A D H F Q C K N G  g746413

443  K C L S K S Q Q C N G K D D C G D G S D E A S C P K - - - -  155179
654  E C V P L V N L C D G H L H C E D G S D E A D C V R F F N G  g746413

469  - - - - - - - - - - - - V N V V - - - T C T K H - T Y R C L N  155179
684  T T N N N G L V R F R I Q S I W H T A C A E N W T T Q I S N  g746413
```

FIGURE 1C

```
484  G L C - - - - L S K G N P E - - - - - C D G - - - - - - - - - 155179
714  D V C Q L L G L G S G N S S K P I F S T D G G P F V K L N T g746413

497  - - - - - - - - - - - K E D C S D G S - - - - - - D E K D C D 155179
744  A P D G H L I L T P S Q Q C L Q D S L I R L Q C N H K S C G g746413

511  C G L R S F T R Q A R V V G G T D A D E G E W P W Q V S L H 155179
774  K K L A A Q D I T P K I V G G S N A K E G A W P W V V G L Y g746413

541  A L G Q G H I C G A S L I S P N W L V S A A H C Y I D D R G 155179
804  - Y G G R L L C G A S L V S S D W L V S A A H C - V Y G R N g746413

571  F R Y S D P T Q W T A F L G L H D Q S Q R S A P G V Q E R R 155179
832  L - - - E P S K W T A I L G L H M K S N L T S P Q T V P R L g746413

601  L K R I I S H P F F N D F T F D Y D I A L L E L E K P A E Y 155179
859  I D E I V I N P H Y N R R R K D N D I A M M H L E F K V N Y g746413

631  S S M V R P I C L P D A S H V F P A G K A I W V T G W G H T 155179
889  T D Y I Q P I C L P E E N Q V F P P G R N C S I A G W G T V g746413

661  Q Y G G T G A L I L Q K G E I R V I N Q T T C E N L L P Q - 155179
919  V Y Q G T T A N I L Q E A D V P L L S N E R C Q Q Q M P E Y g746413
```

FIGURE 1D

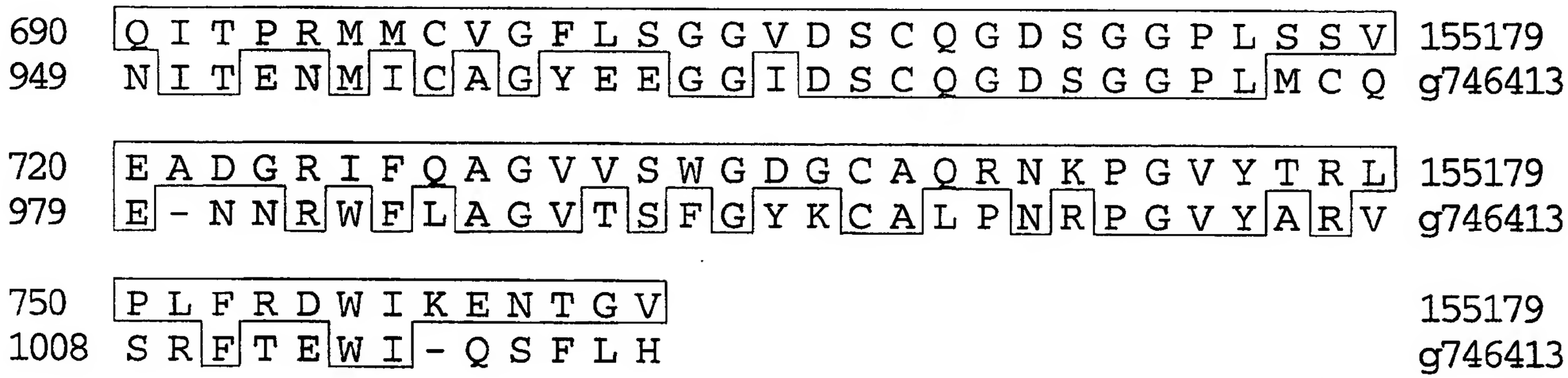
690 Q I T P R M M C V G F L S G G V D S C Q G D S G G P L S S V 155179
949 N I T E N M I C A G Y E E G G I D S C Q G D S G G P L M C Q g746413
720 E A D G R I F Q A G V V S W G D G C A Q R N K P G V Y T R L 155179
979 E - N N R W F L A G V T S F G Y K C A L P N R P G V Y A R V g746413
750 P L F R D W I K E N T G V 155179
1008 S R F T E W I - Q S F L H g746413

FIGURE 1E

```
1    MPAVLGFEGSANKIGVGVV-RDGKVLANPR  2415780
1    M-ICLGLEGTAEKTGVGIVTSDGEVLFNKT  g2826367

30   RTYVTPPGTGFLPGDTARHHRAVILDLLQE  2415780
30    IMY-KPPKQGINPREAADHHAETFPKLIKE g2826367

60   ALTESGLTSQDIDCIAYTKGPGMGAPLVSV  2415780
59   AFE--VVDKNEIDLIAFSQGPGLGPSLRVT  g2826367

90   AVVARTVAQLWNKPLVGVNHCIGHIEMGRL  2415780
87   ATVARTLSLTLKKPIIGVNHCIAHIEIGKL  g2826367

120  ITGATSPTVLYVSGGNTQVIAYSEHRYRIF  2415780
117   TTEAEDPLTLYVSGGNTQVIAYVSKKYRVF g2826367

150  GETIDIAVGNCLDRFARVLKISNDPSPGYN  2415780
147  GETLDIAVGNCLDQFARYVNLPH-PGGPY-  g2826367

180  IEQMAKRGKKLVELPYTVKGMDVSFSGILS  2415780
175  IEELARKGKKLVDLPYTVKGMDIAFSGLLT  g2826367

210  FIEDVAHRMLATGECTPEDLCFSLQETVFA  2415780
205  ----AAMRAYDAGE-RLEDICYSLQEYAFS  g2826367
```

FIGURE 2A

```
240 M L V E I T E R A M A H C G S Q E A L I V G G V G C N V R L 2415780
230 M L T E I T E R A L A H T N K G E V M L V G G V A A N N R L g2826367

270 Q E M M A T M C Q E R G A R L F A T D E R F C I D N G A M I 2415780
260 R E M L K A M C E G Q N V D F Y V P P K E F C G D N G A M I g2826367

300 A Q A G W E M F R A G H R T P L S D S G V T Q R Y R T D E V 2415780
290 A W L G L L M H K N G R W M S L D E T K I I P N Y R T D M V g2826367

330 E V T W R - - - - - - - - - - - - - - - - - - - - - - - - - 2415780
320 E V N W I K E I K G K K R K I P E H L I G K G A E A D I K R g2826367

335 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2415780
350 D S Y L D F D V I I K E R V K K G Y R D E R L D E N I R K S g2826367

335 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2415780
380 R T A R E A R Y L A L V K D F G I P A P Y I F D V D L D N K g2826367

335 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2415780
410 R I M M S Y I N G K L A K D V I E D N L D I A Y K I G E I V g2826367

335 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2415780
440 G K L H K N D V I H N D L T T S N F I F D K D L Y I I D F G g2826367
```

FIGURE 2B

```
335 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  2415780
470 L G K I S N L D E D K A V D L I V F K K A V L S T H H E K F  g2826367

335 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  2415780
500 D E I W E R F L E G Y K S V Y D R W E I I L E L M K D V E R  g2826367

335 - - - - - [D]                                                 2415780
530 R A R Y V E                                                   g2826367
```

FIGURE 2C

```
1    M - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2879274
1    M G R R A L L L L L L S F L A P W A T I A L R P A L R A L G g431321

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2879274
31   S L H L P T N P T S L P A V A K N Y S V L Y F Q Q K V D H F g431321

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2879274
61   G F N T V K T F N Q R Y L V A D K Y W K K N G G S I L F Y T g431321

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2879274
91   G N E G D I I W F C N N T G F M W D V A E E L K A M L V F A g431321

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2879274
121  E H R Y Y G E S L P F G D N S F K D S R H L N F L T S E Q A g431321

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 2879274
151  L A D F A E L I K H L K R T I P G A E N Q P V I A I G G S Y g431321

2    - - - L S A Y L R M K Y P H L V A G A L A A S A P V L A V A 2879274
181  G G M L A A W F R M K Y P H M V V G A L A A S A P I W Q F E g431321

29   G L G D S N Q F F R D V T A D F E G Q S P K C T Q G V R E A 2879274
211  D L V P C G V F M K I V T T D F R K S G P H C S E S I H R S g431321
```

FIGURE 3A

```
59  F R Q I K D L F L Q G A Y D T V R W E F G T - - - C Q P L S  2879274
241 W D A I N R L S N T G S - - G L Q W L T G A L H L C S P L T  g431321

86  D E K D L T Q L F M F A R N A F T V L A M M D Y P Y P T D F  2879274
269 S Q - D I Q H L K D W I S E T W V N L A M V D Y P Y A S N F  g431321

116 L G P L P A N P V K V G C D R L L S E A Q R I T G L R A L A  2879274
298 L Q P L P A W P I K V V C Q - Y L K N P N V S D S L - - L L  g431321

146 G L V Y N A S G S E H C Y D I Y R L Y H S C A D P T G C G T  2879274
325 Q N I F Q A L - - - N V Y Y N Y S G Q V K C L N I S E T A T  g431321

176 G P - D A R A W D Y Q A C T E I N L T F A S N N V T D M F P  2879274
352 S S L G T L G W S Y Q A C T E V V M P F C T N G V D D M F E  g431321

205 D L P F T - D E L R Q R Y C L D T W G V W P R P D W L L T S  2879274
382 P H S W N L K E L S D D - C F Q Q W G V R P R P S W I T T M  g431321

234 F W G G D L R A A S N I I F S N G N L D P W A G G G I R R N  2879274
411 Y G G K N I S S H T N I V F S N G E L D P W S G G G V T K D  g431321

264 L S A S V I A V T I Q G G A H H L D L R A S H P E D P A S V  2879274
441 I T D T L V A V T I S E G A H H L D L R T K N A L D P M S V  g431321
```

FIGURE 3B

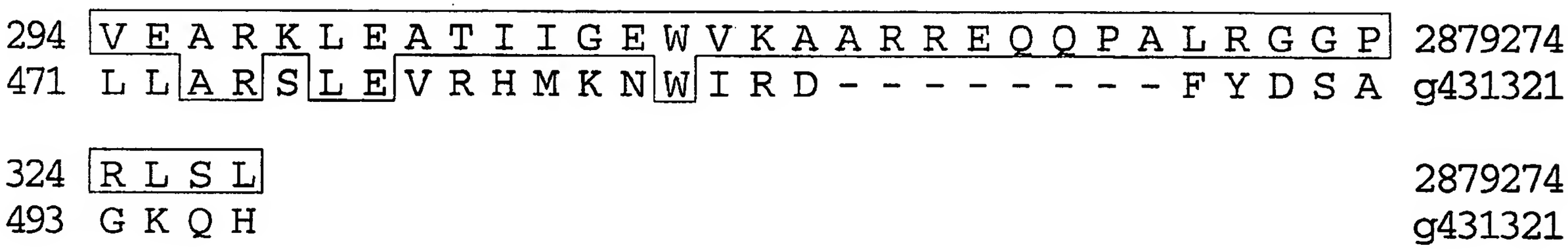
294 V E A R K L E A T I I G E W V K A A R R E Q Q P A L R G G P 2879274
471 L L A R S L E V R H M K N W I R D - - - - - - - - - - F Y D S A g431321
324 R L S L 2879274
493 G K Q H g431321

FIGURE 3C

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
BANDMAN, Olga
HILLMAN, Jennifer L.
TANG, Y. Tom
LAL, Preeti
YUE, Henry
AZIMZAI, Yalda
BAUGHN, Mariah R.
LU, Dyung Aina M.

<120> HUMAN PEPTIDASES

<130> PF-0651 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/228,199; unassigned; 60/132,253; 60/136,653
<151> 1999-01-11; 1999-01-11; 1999-05-03; 1999-05-27

<160> 39

<170> PERL Program

<210> 1
<211> 762
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 155179CD1

<400> 1

```
Met Arg Ile Thr Asn Glu Asn Phe Val Asp Ala Tyr Glu Asn Ser
  1               5                  10                  15
Asn Ser Thr Glu Phe Val Ser Leu Ala Ser Lys Val Lys Asp Ala
             20                  25                  30
Leu Lys Leu Leu Tyr Ser Gly Val Pro Phe Leu Gly Pro Cys His
         35                  40                  45
Lys Glu Ser Ala Val Thr Ala Phe Ser Glu Gly Ser Val Ile Ala
     50                  55                  60
Tyr Tyr Trp Ser Glu Phe Ser Ile Pro Gln His Leu Val Glu Glu
 65                  70                  75
Ala Glu Arg Val Met Ala Glu Glu Arg Val Val Met Leu Pro Pro
             80                  85                  90
Arg Ala Arg Ser Leu Lys Ser Phe Val Val Thr Ser Val Val Ala
         95                 100                 105
Phe Pro Thr Asp Ser Lys Thr Val Gln Arg Thr Gln Asp Asn Ser
    110                 115                 120
Cys Ser Phe Gly Leu His Ala Arg Gly Val Glu Leu Met Arg Phe
125                 130                 135
```

```
Thr Thr Pro Gly Phe Pro Asp Ser Pro Tyr Pro Ala His Ala Arg
                140                 145                 150
Cys Gln Trp Ala Leu Arg Gly Asp Ala Asp Ser Val Leu Ser Leu
                155                 160                 165
Thr Phe Arg Ser Phe Asp Leu Ala Ser Cys Asp Glu Arg Gly Ser
                170                 175                 180
Asp Leu Val Thr Val Tyr Asn Thr Leu Ser Pro Met Glu Pro His
                185                 190                 195
Ala Leu Val Gln Leu Cys Gly Thr Tyr Pro Pro Ser Tyr Asn Leu
                200                 205                 210
Thr Phe His Ser Ser Gln Asn Val Leu Leu Ile Thr Leu Ile Thr
                215                 220                 225
Asn Thr Glu Arg Arg His Pro Gly Phe Glu Ala Thr Phe Phe Gln
                230                 235                 240
Leu Pro Arg Met Ser Ser Cys Gly Gly Arg Leu Arg Lys Ala Gln
                245                 250                 255
Gly Thr Phe Asn Ser Pro Tyr Tyr Pro Gly His Tyr Pro Pro Asn
                260                 265                 270
Ile Asp Cys Thr Trp Asn Ile Glu Val Pro Asn Asn Gln His Val
                275                 280                 285
Lys Val Arg Phe Lys Phe Phe Tyr Leu Leu Glu Pro Gly Val Pro
                290                 295                 300
Ala Gly Thr Cys Pro Lys Asp Tyr Val Glu Ile Asn Gly Glu Lys
                305                 310                 315
Tyr Cys Gly Glu Arg Ser Gln Phe Val Val Thr Ser Asn Ser Asn
                320                 325                 330
Lys Ile Thr Val Arg Phe His Ser Asp Gln Ser Tyr Thr Asp Thr
                335                 340                 345
Gly Phe Leu Ala Glu Tyr Leu Ser Tyr Asp Ser Ser Asp Pro Cys
                350                 355                 360
Pro Gly Gln Phe Thr Cys Arg Thr Gly Arg Cys Ile Arg Lys Glu
                365                 370                 375
Leu Arg Cys Asp Gly Trp Ala Asp Cys Thr Asp His Ser Asp Glu
                380                 385                 390
Leu Asn Cys Ser Cys Asp Ala Gly His Gln Phe Thr Cys Lys Asn
                395                 400                 405
Lys Phe Cys Lys Pro Leu Phe Trp Val Cys Asp Ser Val Asn Asp
                410                 415                 420
Cys Gly Asp Asn Ser Asp Glu Gln Gly Cys Ser Cys Pro Ala Gln
                425                 430                 435
Thr Phe Arg Cys Ser Asn Gly Lys Cys Leu Ser Lys Ser Gln Gln
                440                 445                 450
Cys Asn Gly Lys Asp Asp Cys Gly Asp Gly Ser Asp Glu Ala Ser
                455                 460                 465
Cys Pro Lys Val Asn Val Val Thr Cys Thr Lys His Thr Tyr Arg
                470                 475                 480
Cys Leu Asn Gly Leu Cys Leu Ser Lys Gly Asn Pro Glu Cys Asp
                485                 490                 495
Gly Lys Glu Asp Cys Ser Asp Gly Ser Asp Glu Lys Asp Cys Asp
                500                 505                 510
Cys Gly Leu Arg Ser Phe Thr Arg Gln Ala Arg Val Val Gly Gly
                515                 520                 525
Thr Asp Ala Asp Glu Gly Glu Trp Pro Trp Gln Val Ser Leu His
                530                 535                 540
Ala Leu Gly Gln Gly His Ile Cys Gly Ala Ser Leu Ile Ser Pro
```

```
                545                 550                 555
Asn Trp Leu Val Ser Ala Ala His Cys Tyr Ile Asp Asp Arg Gly
                560                 565                 570
Phe Arg Tyr Ser Asp Pro Thr Gln Trp Thr Ala Phe Leu Gly Leu
                575                 580                 585
His Asp Gln Ser Gln Arg Ser Ala Pro Gly Val Gln Glu Arg Arg
                590                 595                 600
Leu Lys Arg Ile Ile Ser His Pro Phe Phe Asn Asp Phe Thr Phe
                605                 610                 615
Asp Tyr Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro Ala Glu Tyr
                620                 625                 630
Ser Ser Met Val Arg Pro Ile Cys Leu Pro Asp Ala Ser His Val
                635                 640                 645
Phe Pro Ala Gly Lys Ala Ile Trp Val Thr Gly Trp Gly His Thr
                650                 655                 660
Gln Tyr Gly Gly Thr Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile
                665                 670                 675
Arg Val Ile Asn Gln Thr Thr Cys Glu Asn Leu Leu Pro Gln Gln
                680                 685                 690
Ile Thr Pro Arg Met Met Cys Val Gly Phe Leu Ser Gly Gly Val
                695                 700                 705
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Ser Ser Val Glu
                710                 715                 720
Ala Asp Gly Arg Ile Phe Gln Ala Gly Val Val Ser Trp Gly Asp
                725                 730                 735
Gly Cys Ala Gln Arg Asn Lys Pro Gly Val Tyr Thr Arg Leu Pro
                740                 745                 750
Leu Phe Arg Asp Trp Ile Lys Glu Asn Thr Gly Val
                755                 760
```

<210> 2
<211> 335
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2415780CD1

<400> 2

```
Met Pro Ala Val Leu Gly Phe Glu Gly Ser Ala Asn Lys Ile Gly
  1                 5                  10                  15
Val Gly Val Val Arg Asp Gly Lys Val Leu Ala Asn Pro Arg Arg
                 20                  25                  30
Thr Tyr Val Thr Pro Pro Gly Thr Gly Phe Leu Pro Gly Asp Thr
                 35                  40                  45
Ala Arg His His Arg Ala Val Ile Leu Asp Leu Leu Gln Glu Ala
                 50                  55                  60
Leu Thr Glu Ser Gly Leu Thr Ser Gln Asp Ile Asp Cys Ile Ala
                 65                  70                  75
Tyr Thr Lys Gly Pro Gly Met Gly Ala Pro Leu Val Ser Val Ala
                 80                  85                  90
Val Val Ala Arg Thr Val Ala Gln Leu Trp Asn Lys Pro Leu Val
                 95                 100                 105
```

```
Gly Val Asn His Cys Ile Gly His Ile Glu Met Gly Arg Leu Ile
                110                 115                 120
Thr Gly Ala Thr Ser Pro Thr Val Leu Tyr Val Ser Gly Gly Asn
                125                 130                 135
Thr Gln Val Ile Ala Tyr Ser Glu His Arg Tyr Arg Ile Phe Gly
                140                 145                 150
Glu Thr Ile Asp Ile Ala Val Gly Asn Cys Leu Asp Arg Phe Ala
                155                 160                 165
Arg Val Leu Lys Ile Ser Asn Asp Pro Ser Pro Gly Tyr Asn Ile
                170                 175                 180
Glu Gln Met Ala Lys Arg Gly Lys Lys Leu Val Glu Leu Pro Tyr
                185                 190                 195
Thr Val Lys Gly Met Asp Val Ser Phe Ser Gly Ile Leu Ser Phe
                200                 205                 210
Ile Glu Asp Val Ala His Arg Met Leu Ala Thr Gly Glu Cys Thr
                215                 220                 225
Pro Glu Asp Leu Cys Phe Ser Leu Gln Glu Thr Val Phe Ala Met
                230                 235                 240
Leu Val Glu Ile Thr Glu Arg Ala Met Ala His Cys Gly Ser Gln
                245                 250                 255
Glu Ala Leu Ile Val Gly Gly Val Gly Cys Asn Val Arg Leu Gln
                260                 265                 270
Glu Met Met Ala Thr Met Cys Gln Glu Arg Gly Ala Arg Leu Phe
                275                 280                 285
Ala Thr Asp Glu Arg Phe Cys Ile Asp Asn Gly Ala Met Ile Ala
                290                 295                 300
Gln Ala Gly Trp Glu Met Phe Arg Ala Gly His Arg Thr Pro Leu
                305                 310                 315
Ser Asp Ser Gly Val Thr Gln Arg Tyr Arg Thr Asp Glu Val Glu
                320                 325                 330
Val Thr Trp Arg Asp
                335
```

<210> 3
<211> 327
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2879274CD1

<400> 3

```
Met Leu Ser Ala Tyr Leu Arg Met Lys Tyr Pro His Leu Val Ala
  1               5                  10                  15
Gly Ala Leu Ala Ala Ser Ala Pro Val Leu Ala Val Ala Gly Leu
                20                  25                  30
Gly Asp Ser Asn Gln Phe Phe Arg Asp Val Thr Ala Asp Phe Glu
                35                  40                  45
Gly Gln Ser Pro Lys Cys Thr Gln Gly Val Arg Glu Ala Phe Arg
                50                  55                  60
Gln Ile Lys Asp Leu Phe Leu Gln Gly Ala Tyr Asp Thr Val Arg
                65                  70                  75
Trp Glu Phe Gly Thr Cys Gln Pro Leu Ser Asp Glu Lys Asp Leu
```

```
                80                  85                  90
Thr Gln Leu Phe Met Phe Ala Arg Asn Ala Phe Thr Val Leu Ala
                95                 100                 105
Met Met Asp Tyr Pro Tyr Pro Thr Asp Phe Leu Gly Pro Leu Pro
               110                 115                 120
Ala Asn Pro Val Lys Val Gly Cys Asp Arg Leu Leu Ser Glu Ala
               125                 130                 135
Gln Arg Ile Thr Gly Leu Arg Ala Leu Ala Gly Leu Val Tyr Asn
               140                 145                 150
Ala Ser Gly Ser Glu His Cys Tyr Asp Ile Tyr Arg Leu Tyr His
               155                 160                 165
Ser Cys Ala Asp Pro Thr Gly Cys Gly Thr Gly Pro Asp Ala Arg
               170                 175                 180
Ala Trp Asp Tyr Gln Ala Cys Thr Glu Ile Asn Leu Thr Phe Ala
               185                 190                 195
Ser Asn Asn Val Thr Asp Met Phe Pro Asp Leu Pro Phe Thr Asp
               200                 205                 210
Glu Leu Arg Gln Arg Tyr Cys Leu Asp Thr Trp Gly Val Trp Pro
               215                 220                 225
Arg Pro Asp Trp Leu Leu Thr Ser Phe Trp Gly Gly Asp Leu Arg
               230                 235                 240
Ala Ala Ser Asn Ile Ile Phe Ser Asn Gly Asn Leu Asp Pro Trp
               245                 250                 255
Ala Gly Gly Gly Ile Arg Arg Asn Leu Ser Ala Ser Val Ile Ala
               260                 265                 270
Val Thr Ile Gln Gly Gly Ala His His Leu Asp Leu Arg Ala Ser
               275                 280                 285
His Pro Glu Asp Pro Ala Ser Val Val Glu Ala Arg Lys Leu Glu
               290                 295                 300
Ala Thr Ile Ile Gly Glu Trp Val Lys Ala Ala Arg Arg Glu Gln
               305                 310                 315
Gln Pro Ala Leu Arg Gly Gly Pro Arg Leu Ser Leu
               320                 325
```

<210> 4
<211> 471
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 358050CD1

<400> 4

```
Met Ala Ala Met Glu Thr Glu Thr Ala Pro Leu Thr Leu Glu Ser
  1               5                  10                  15
Leu Pro Thr Asp Pro Leu Leu Leu Ile Leu Ser Phe Leu Asp Tyr
                20                  25                  30
Arg Asp Leu Ile Asn Cys Cys Tyr Val Ser Arg Arg Leu Ser Gln
                35                  40                  45
Leu Ser Ser His Asp Pro Leu Trp Arg Arg His Cys Lys Lys Tyr
                50                  55                  60
Trp Leu Ile Ser Glu Glu Glu Lys Thr Gln Lys Asn Gln Cys Trp
                65                  70                  75
```

```
Lys Ser Leu Phe Ile Asp Thr Tyr Ser Asp Val Gly Arg Tyr Ile
                80                  85                  90
Asp His Tyr Ala Ala Ile Lys Lys Ala Trp Asp Asp Leu Lys Lys
                95                 100                 105
Tyr Leu Glu Pro Arg Cys Pro Arg Met Val Leu Ser Leu Lys Glu
               110                 115                 120
Gly Ala Arg Glu Glu Asp Leu Asp Ala Val Glu Ala Gln Ile Gly
               125                 130                 135
Cys Lys Leu Pro Asp Asp Tyr Arg Cys Ser Tyr Arg Ile His Asn
               140                 145                 150
Gly Gln Lys Leu Val Val Pro Gly Leu Leu Gly Ser Met Ala Leu
               155                 160                 165
Ser Asn His Tyr Arg Ser Glu Asp Leu Leu Asp Val Asp Thr Ala
               170                 175                 180
Ala Gly Gly Phe Gln Gln Arg Gln Gly Leu Lys Tyr Cys Leu Pro
               185                 190                 195
Leu Thr Phe Cys Ile His Thr Gly Leu Ser Gln Tyr Ile Ala Val
               200                 205                 210
Glu Ala Ala Glu Gly Arg Asn Lys Asn Glu Val Phe Tyr Gln Cys
               215                 220                 225
Pro Asp Gln Met Ala Arg Asn Pro Ala Ala Ile Asp Met Phe Ile
               230                 235                 240
Ile Gly Ala Thr Phe Thr Asp Trp Phe Thr Ser Tyr Val Lys Asn
               245                 250                 255
Val Val Ser Gly Gly Phe Pro Ile Ile Arg Asp Gln Ile Phe Arg
               260                 265                 270
Tyr Val His Asp Pro Glu Cys Val Ala Thr Thr Gly Asp Ile Thr
               275                 280                 285
Val Ser Val Ser Thr Ser Phe Leu Pro Glu Leu Ser Ser Val His
               290                 295                 300
Pro Pro His Tyr Phe Phe Thr Tyr Arg Ile Arg Ile Glu Met Ser
               305                 310                 315
Lys Asp Ala Leu Pro Glu Lys Ala Cys Gln Leu Asp Ser Arg Tyr
               320                 325                 330
Trp Arg Ile Thr Asn Ala Lys Gly Asp Val Glu Glu Val Gln Gly
               335                 340                 345
Pro Gly Val Val Gly Glu Phe Pro Ile Ile Ser Pro Gly Arg Val
               350                 355                 360
Tyr Glu Tyr Thr Ser Cys Thr Thr Phe Ser Thr Thr Ser Gly Tyr
               365                 370                 375
Met Glu Gly Tyr Tyr Thr Phe His Phe Leu Tyr Phe Lys Asp Lys
               380                 385                 390
Ile Phe Asn Val Ala Ile Pro Arg Phe His Met Ala Cys Pro Thr
               395                 400                 405
Phe Arg Val Ser Ile Ala Arg Leu Glu Met Gly Pro Asp Glu Tyr
               410                 415                 420
Glu Glu Met Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Glu
               425                 430                 435
Asp Asp Asp Ser Ala Asp Met Asp Glu Ser Asp Glu Asp Asp Glu
               440                 445                 450
Glu Glu Arg Arg Arg Arg Val Phe Asp Val Pro Ile Arg Arg Arg
               455                 460                 465
Arg Cys Ser Arg Leu Phe
               470
```

<210> 5
<211> 60
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 700745CD1

<400> 5
Met Thr Pro Trp Leu Gly Leu Ile Val Leu Leu Gly Ser Trp Ser
1 5 10 15
Leu Gly Asp Trp Gly Ala Glu Ala Cys Thr Cys Ser Pro Ser His
20 25 30
Pro Gln Asp Ala Phe Cys Asn Ser Asp Ile Gly Lys Arg Ser Trp
35 40 45
Cys Pro Ala Arg Ala Pro Arg Cys Ser Gln Asp Cys Ser Ala Ala
50 55 60

<210> 6
<211> 399
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2026480CD1

<400> 6
Met Ala His Ile Thr Ile Asn Gln Tyr Leu Gln Gln Val Tyr Glu
1 5 10 15
Ala Ile Asp Ser Arg Asp Gly Ala Ser Cys Ala Glu Leu Val Ser
20 25 30
Phe Lys His Pro His Val Ala Asn Pro Arg Leu Gln Met Ala Ser
35 40 45
Pro Glu Glu Lys Cys Gln Gln Val Leu Glu Pro Pro Tyr Asp Glu
50 55 60
Met Phe Ala Ala His Leu Arg Cys Thr Tyr Ala Val Gly Asn His
65 70 75
Asp Phe Ile Glu Ala Tyr Lys Cys Gln Thr Val Ile Val Gln Ser
80 85 90
Phe Leu Arg Ala Phe Gln Ala His Lys Glu Glu Asn Trp Ala Leu
95 100 105
Pro Val Met Tyr Ala Val Ala Leu Asp Leu Arg Val Phe Ala Asn
110 115 120
Asn Ala Asp Gln Gln Leu Val Lys Lys Gly Lys Ser Lys Val Gly
125 130 135
Asp Met Leu Glu Lys Ala Ala Glu Leu Leu Met Ser Cys Phe Arg
140 145 150
Val Cys Ala Ser Asp Thr Arg Ala Gly Ile Glu Asp Ser Lys Lys
155 160 165
Trp Gly Met Leu Phe Leu Val Asn Gln Leu Phe Lys Ile Tyr Phe
170 175 180
Lys Ile Asn Lys Leu His Leu Cys Lys Pro Leu Ile Arg Ala Ile

185 190 195
Asp Ser Ser Asn Leu Lys Asp Asp Tyr Ser Thr Ala Gln Arg Val
200 205 210
Thr Tyr Lys Tyr Tyr Val Gly Arg Lys Ala Met Phe Asp Ser Asp
215 220 225
Phe Lys Gln Ala Glu Glu Tyr Leu Ser Phe Ala Phe Glu His Cys
230 235 240
His Arg Ser Ser Gln Lys Asn Lys Arg Met Ile Leu Ile Tyr Leu
245 250 255
Leu Pro Val Lys Met Leu Leu Gly His Met Pro Thr Val Glu Leu
260 265 270
Leu Lys Lys Tyr His Leu Met Gln Phe Ala Glu Val Thr Arg Ala
275 280 285
Val Ser Glu Gly Asn Leu Leu Leu Leu His Glu Ala Leu Ala Lys
290 295 300
His Glu Ala Phe Phe Ile Arg Cys Gly Ile Phe Leu Ile Leu Glu
305 310 315
Lys Leu Lys Ile Ile Thr Tyr Arg Asn Leu Phe Lys Lys Val Tyr
320 325 330
Leu Leu Leu Lys Thr His Gln Leu Ser Leu Asp Ala Phe Leu Val
335 340 345
Ala Leu Lys Phe Met Gln Val Glu Asp Val Asp Ile Asp Glu Val
350 355 360
Gln Cys Ile Leu Ala Asn Leu Ile Tyr Met Gly His Val Lys Gly
365 370 375
Tyr Ile Ser His Gln His Gln Lys Leu Val Val Ser Lys Gln Asn
380 385 390
Pro Phe Pro Pro Leu Ser Thr Val Cys
395

<210> 7
<211> 106
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2132401CD1

<400> 7
Met Ile Glu Glu Lys Ser Asp Ile Glu Thr Leu Asp Ile Pro Glu
1 5 10 15
Pro Pro Pro Asn Ser Gly Tyr Glu Cys Gln Leu Arg Leu Arg Leu
20 25 30
Ser Thr Gly Lys Asp Leu Lys Leu Val Val Arg Ser Thr Asp Thr
35 40 45
Val Phe His Met Lys Arg Arg Leu His Ala Ala Glu Gly Val Glu
50 55 60
Pro Gly Ser Gln Arg Trp Phe Phe Ser Gly Arg Pro Leu Thr Asp
65 70 75
Lys Met Lys Phe Glu Glu Leu Lys Ile Pro Lys Asp Tyr Val Val
80 85 90
Gln Val Ile Val Ser Gln Pro Val Gln Asn Pro Thr Pro Val Glu
95 100 105

Asn

<210> 8
<211> 267
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2568875CD1

<400> 8

Met	Ser	Asp	Glu	Asp	Ser	Cys	Val	Ala	Cys	Gly	Ser	Leu	Arg	Thr	
1				5				10						15	
Ala	Gly	Pro	Gln	Ala	Gly	Ala	Pro	Ser	Pro	Trp	Pro	Trp	Glu	Ala	
				20				25						30	
Arg	Leu	Met	His	Gln	Gly	Gln	Leu	Ala	Cys	Gly	Gly	Ala	Leu	Val	
				35				40						45	
Ser	Glu	Glu	Ala	Val	Leu	Thr	Ala	Ala	His	Cys	Phe	Ile	Gly	Arg	
				50				55						60	
Gln	Ala	Pro	Glu	Glu	Trp	Ser	Val	Gly	Leu	Gly	Thr	Arg	Pro	Glu	
				65				70						75	
Glu	Trp	Gly	Leu	Lys	Gln	Leu	Ile	Leu	His	Gly	Ala	Tyr	Thr	His	
				80				85						90	
Pro	Glu	Gly	Gly	Tyr	Asp	Met	Ala	Leu	Leu	Leu	Leu	Ala	Gln	Pro	
				95				100						105	
Val	Thr	Leu	Gly	Ala	Ser	Leu	Arg	Pro	Leu	Cys	Leu	Pro	Tyr	Ala	
				110				115						120	
Asp	His	His	Leu	Pro	Asp	Gly	Glu	Arg	Gly	Trp	Val	Leu	Gly	Arg	
				125				130						135	
Ala	Arg	Pro	Gly	Ala	Gly	Ile	Ser	Ser	Leu	Gln	Thr	Val	Pro	Val	
				140				145						150	
Thr	Leu	Leu	Gly	Pro	Arg	Ala	Cys	Ser	Arg	Leu	His	Ala	Ala	Pro	
				155				160						165	
Gly	Gly	Asp	Gly	Ser	Pro	Ile	Leu	Pro	Gly	Met	Val	Cys	Thr	Ser	
				170				175						180	
Ala	Val	Gly	Glu	Leu	Pro	Ser	Cys	Glu	Gly	Leu	Ser	Gly	Ala	Pro	
				185				190						195	
Leu	Val	His	Glu	Val	Arg	Gly	Thr	Trp	Phe	Leu	Ala	Gly	Leu	His	
				200				205						210	
Ser	Phe	Gly	Asp	Ala	Cys	Gln	Gly	Pro	Ala	Arg	Pro	Ala	Val	Phe	
				215				220						225	
Thr	Ala	Leu	Pro	Ala	Tyr	Glu	Asp	Trp	Val	Ser	Ser	Leu	Asp	Trp	
				230				235						240	
Gln	Val	Tyr	Phe	Ala	Glu	Glu	Pro	Glu	Pro	Glu	Ala	Glu	Pro	Gly	
				245				250						255	
Ser	Cys	Leu	Ala	Asn	Ile	Ser	Gln	Pro	Thr	Ser	Cys				
				260				265							

<210> 9
<211> 123
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3408908CD1

<400> 9
Met Arg Thr Gln Ser Leu Leu Leu Leu Gly Ala Leu Leu Ala Val
1 5 10 15
Gly Ser Gln Leu Pro Ala Val Phe Gly Arg Lys Lys Gly Glu Lys
20 25 30
Ser Gly Gly Cys Pro Pro Asp Asp Gly Pro Cys Leu Leu Ser Val
35 40 45
Pro Asp Gln Cys Val Glu Asp Ser Gln Cys Pro Leu Thr Arg Lys
50 55 60
Cys Cys Tyr Arg Ala Cys Phe Arg Gln Cys Val Pro Arg Val Ser
65 70 75
Val Lys Leu Gly Ser Cys Pro Glu Asp Gln Leu Arg Cys Leu Ser
80 85 90
Pro Met Asn His Leu Cys Tyr Lys Asp Ser Asp Cys Ser Gly Lys
95 100 105
Lys Arg Cys Cys His Ser Ala Cys Gly Arg Asp Cys Arg Asp Pro
110 115 120
Ala Arg Gly

<210> 10
<211> 513
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3772696CD1

<400> 10
Met Lys Arg Leu Leu Leu Leu Cys Leu Phe Phe Ile Thr Phe Ser
1 5 10 15
Ser Ala Phe Pro Leu Val Arg Met Thr Glu Asn Glu Glu Asn Met
20 25 30
Gln Leu Ala Gln Ala Tyr Leu Asn Gln Phe Tyr Ser Leu Glu Ile
35 40 45
Glu Gly Asn His Leu Val Gln Ser Lys Asn Arg Ser Leu Ile Asp
50 55 60
Asp Lys Ile Arg Glu Met Gln Ala Phe Phe Gly Leu Thr Val Thr
65 70 75
Gly Lys Leu Asp Ser Asn Thr Leu Glu Ile Met Lys Thr Pro Arg
80 85 90
Cys Gly Val Pro Asp Val Gly Gln Tyr Gly Tyr Thr Leu Pro Gly
95 100 105
Trp Arg Lys Tyr Asn Leu Thr Tyr Arg Ile Ile Asn Tyr Thr Pro
110 115 120
Asp Met Ala Arg Ala Ala Val Asp Glu Ala Ile Gln Glu Gly Leu
125 130 135
Glu Val Trp Ser Lys Val Thr Pro Leu Lys Phe Thr Lys Ile Ser
140 145 150
Lys Gly Ile Ala Asp Ile Met Ile Ala Phe Arg Thr Arg Val His

```
                155                 160                 165
Gly Arg Cys Pro Arg Tyr Phe Asp Gly Pro Leu Gly Val Leu Gly
                170                 175                 180
His Ala Phe Pro Pro Gly Pro Gly Leu Gly Gly Asp Thr His Phe
                185                 190                 195
Asp Glu Asp Glu Asn Trp Thr Lys Asp Gly Ala Gly Phe Asn Leu
                200                 205                 210
Phe Leu Val Ala Ala His Glu Phe Gly His Ala Leu Gly Leu Ser
                215                 220                 225
His Ser Asn Asp Gln Thr Ala Leu Met Phe Pro Asn Tyr Val Ser
                230                 235                 240
Leu Asp Pro Arg Lys Tyr Pro Leu Ser Gln Asp Asp Ile Asn Gly
                245                 250                 255
Ile Gln Ser Ile Tyr Gly Gly Leu Pro Lys Val Pro Ala Lys Pro
                260                 265                 270
Lys Glu Pro Thr Ile Pro His Ala Cys Asp Pro Asp Leu Thr Phe
                275                 280                 285
Asp Ala Ile Thr Thr Phe Arg Arg Glu Val Met Phe Phe Lys Gly
                290                 295                 300
Arg His Leu Trp Arg Ile Tyr Tyr Asp Ile Thr Asp Val Glu Phe
                305                 310                 315
Glu Leu Ile Ala Ser Phe Trp Pro Ser Leu Pro Ala Asp Leu Gln
                320                 325                 330
Ala Ala Tyr Glu Asn Pro Arg Asp Lys Ile Leu Val Phe Lys Asp
                335                 340                 345
Glu Asn Phe Trp Met Ile Arg Gly Tyr Ala Val Leu Pro Asp Tyr
                350                 355                 360
Pro Lys Ser Ile His Thr Leu Gly Phe Pro Gly Arg Val Lys Lys
                365                 370                 375
Ile Asp Ala Ala Val Cys Asp Lys Thr Thr Arg Lys Thr Tyr Phe
                380                 385                 390
Phe Val Gly Ile Trp Cys Trp Arg Phe Asp Glu Met Thr Gln Thr
                395                 400                 405
Met Asp Lys Gly Phe Pro Gln Arg Val Val Lys His Phe Pro Gly
                410                 415                 420
Ile Ser Ile Arg Val Asp Ala Ala Phe Gln Tyr Lys Gly Phe Phe
                425                 430                 435
Phe Phe Ser Arg Gly Ser Lys Gln Phe Glu Tyr Asn Ile Lys Thr
                440                 445                 450
Lys Asn Ile Thr Arg Ile Met Arg Thr Asn Thr Trp Phe Gln Cys
                455                 460                 465
Lys Glu Pro Lys Asn Ser Ser Phe Gly Phe Asp Ile Asn Lys Glu
                470                 475                 480
Lys Ala His Ser Gly Gly Ile Lys Ile Leu Tyr His Lys Ser Leu
                485                 490                 495
Ser Leu Phe Ile Phe Gly Ile Val His Leu Leu Lys Asn Thr Ser
                500                 505                 510
Ile Tyr Gln
```

<210> 11
<211> 326
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5388674CD1

<400> 11

```
Met Lys Pro Ser Ser Gln Pro Val Ile Ser Leu Asp Pro Leu Pro
 1               5                  10                  15
Cys Ile Leu His Gln Ile Gly Ser Pro Pro Thr Leu Arg Leu Pro
            20                  25                  30
Lys Thr Leu Asn Ser Ser Ser Val Ile Leu Thr Glu Arg His Pro
        35                  40                  45
Leu Gln Thr Asn Ala Ala Phe Ile Tyr Ser Pro Leu Val Asn Thr
    50                  55                  60
Gly Ser Leu Gly Asn Thr Arg Ile Ile Ser Glu Glu Tyr Ile Lys
65                  70                  75
Trp Leu Thr Gly Tyr Cys Lys Ala Tyr Phe Tyr Gly Leu Arg Val
                80                  85                  90
Lys Leu Leu Glu Pro Val Pro Val Ser Val Thr Arg Cys Ser Phe
            95                  100                 105
Arg Val Asn Glu Asn Thr His Asn Leu Gln Ile His Ala Gly Asp
        110                 115                 120
Ile Leu Lys Phe Leu Lys Lys Lys Lys Pro Glu Asp Ala Phe Cys
    125                 130                 135
Val Val Gly Ile Thr Met Ile Asp Leu Tyr Pro Arg Asp Ser Trp
140                 145                 150
Asn Phe Val Phe Gly Gln Ala Ser Leu Thr Asp Gly Val Gly Ile
155                 160                 165
Phe Ser Phe Ala Arg Tyr Gly Ser Asp Phe Tyr Ser Met His Tyr
                170                 175                 180
Lys Gly Lys Val Lys Lys Leu Lys Lys Thr Ser Ser Ser Asp Tyr
            185                 190                 195
Ser Ile Phe Asp Asn Tyr Tyr Ile Pro Glu Ile Thr Ser Val Leu
        200                 205                 210
Leu Leu Arg Ser Cys Lys Thr Leu Thr His Glu Ile Gly His Ile
    215                 220                 225
Phe Gly Leu Arg His Cys Gln Trp Leu Ala Cys Leu Met Gln Gly
230                 235                 240
Ser Asn His Leu Glu Glu Ala Asp Arg Arg Pro Leu Asn Leu Cys
245                 250                 255
Pro Ile Cys Leu His Lys Leu Gln Cys Ala Val Gly Phe Ser Ile
            260                 265                 270
Val Glu Arg Tyr Lys Ala Leu Val Arg Trp Ile Asp Asp Glu Ser
        275                 280                 285
Ser Asp Thr Pro Gly Ala Thr Pro Glu His Ser His Glu Asp Asn
    290                 295                 300
Gly Asn Leu Pro Lys Pro Val Glu Ala Phe Lys Glu Trp Lys Glu
305                 310                 315
Trp Ile Ile Lys Cys Leu Ala Val Leu Gln Lys
            320                 325
```

<210> 12
<211> 823
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1873102CD1

<400> 12

```
Met Gly Lys Lys Arg Thr Lys Gly Lys Thr Val Pro Ile Asp Asp
 1               5                  10                  15
Ser Ser Glu Thr Leu Glu Pro Val Cys Arg His Ile Arg Lys Gly
            20                  25                  30
Leu Glu Gln Gly Asn Leu Lys Lys Ala Leu Val Asn Val Glu Trp
            35                  40                  45
Asn Ile Cys Gln Asp Cys Lys Thr Asp Asn Lys Val Lys Asp Lys
            50                  55                  60
Ala Glu Glu Glu Thr Glu Glu Lys Pro Ser Val Trp Leu Cys Leu
            65                  70                  75
Lys Cys Gly His Gln Gly Cys Gly Arg Asn Ser Gln Glu Gln His
            80                  85                  90
Ala Leu Lys His Tyr Leu Thr Pro Arg Ser Glu Pro His Cys Leu
            95                 100                 105
Val Leu Ser Leu Asp Asn Trp Ser Val Trp Phe Tyr Val Cys Asp
           110                 115                 120
Asn Glu Val Gln Tyr Cys Ser Ser Asn Gln Leu Gly Gln Val Val
           125                 130                 135
Asp Tyr Val Arg Lys Gln Ala Ser Ile Thr Thr Pro Lys Pro Ala
           140                 145                 150
Glu Lys Asp Asn Gly Asn Ile Glu Leu Glu Asn Lys Lys Leu Glu
           155                 160                 165
Lys Glu Ser Lys Asn Glu Gln Glu Arg Glu Lys Lys Glu Asn Met
           170                 175                 180
Ala Lys Glu Asn Pro Pro Met Asn Ser Pro Cys Gln Ile Thr Val
           185                 190                 195
Lys Gly Leu Ser Asn Leu Gly Asn Thr Cys Phe Phe Asn Ala Val
           200                 205                 210
Met Gln Asn Leu Ser Gln Thr Pro Val Leu Arg Glu Leu Leu Lys
           215                 220                 225
Glu Val Lys Met Ser Gly Thr Ile Val Lys Ile Glu Pro Pro Asp
           230                 235                 240
Leu Ala Leu Thr Glu Pro Leu Glu Ile Asn Leu Glu Pro Pro Gly
           245                 250                 255
Pro Leu Thr Leu Ala Met Ser Gln Phe Leu Asn Glu Met Gln Glu
           260                 265                 270
Thr Lys Lys Gly Val Val Thr Pro Lys Glu Leu Phe Ser Gln Val
           275                 280                 285
Cys Lys Lys Ala Val Arg Phe Lys Gly Tyr Gln Gln Gln Asp Ser
           290                 295                 300
Gln Glu Leu Leu Arg Tyr Leu Leu Asp Gly Met Arg Ala Glu Glu
           305                 310                 315
His Gln Arg Val Ser Lys Gly Ile Leu Lys Ala Phe Gly Asn Ser
           320                 325                 330
Thr Glu Lys Leu Asp Glu Glu Leu Lys Asn Lys Val Lys Asp Tyr
           335                 340                 345
Glu Lys Lys Lys Ser Met Pro Ser Phe Val Asp Arg Ile Phe Gly
           350                 355                 360
Gly Glu Leu Thr Ser Met Ile Met Cys Asp Gln Cys Arg Thr Val
           365                 370                 375
```

```
Ser Leu Val His Glu Ser Phe Leu Asp Leu Ser Leu Pro Val Leu
                380                 385                 390
Asp Asp Gln Ser Gly Lys Lys Ser Val Asn Asp Lys Asn Leu Lys
                395                 400                 405
Lys Thr Val Glu Asp Glu Asp Gln Asp Ser Glu Glu Glu Lys Asp
                410                 415                 420
Asn Asp Ser Tyr Ile Lys Glu Arg Ser Asp Ile Pro Ser Gly Thr
                425                 430                 435
Ser Lys His Leu Gln Lys Lys Ala Lys Lys Gln Ala Lys Lys Gln
                440                 445                 450
Ala Lys Asn Gln Arg Arg Gln Gln Lys Ile Gln Gly Lys Val Leu
                455                 460                 465
His Leu Asn Asp Ile Cys Thr Ile Asp His Pro Glu Asp Ser Glu
                470                 475                 480
Tyr Glu Ala Glu Met Ser Leu Gln Gly Glu Val Asn Ile Lys Ser
                485                 490                 495
Asn His Ile Ser Gln Glu Gly Val Met His Lys Glu Tyr Cys Val
                500                 505                 510
Asn Gln Lys Asp Leu Asn Gly Gln Ala Lys Met Ile Glu Ser Val
                515                 520                 525
Thr Asp Asn Gln Lys Ser Thr Glu Glu Val Asp Met Lys Asn Ile
                530                 535                 540
Asn Met Asp Asn Asp Leu Glu Val Leu Thr Ser Ser Pro Thr Arg
                545                 550                 555
Asn Leu Asn Gly Ala Tyr Leu Thr Glu Gly Ser Asn Gly Glu Val
                560                 565                 570
Asp Ile Ser Asn Gly Phe Lys Asn Leu Asn Leu Asn Ala Ala Leu
                575                 580                 585
His Pro Asp Glu Ile Asn Ile Glu Ile Leu Asn Asp Ser His Thr
                590                 595                 600
Pro Gly Thr Lys Val Tyr Glu Val Val Asn Glu Asp Pro Glu Thr
                605                 610                 615
Ala Phe Cys Thr Leu Ala Asn Arg Glu Val Phe Asn Thr Asp Glu
                620                 625                 630
Cys Ser Ile Gln His Cys Leu Tyr Gln Phe Thr Arg Asn Glu Lys
                635                 640                 645
Leu Arg Asp Ala Asn Lys Leu Leu Cys Glu Val Cys Thr Arg Arg
                650                 655                 660
Gln Cys Asn Gly Pro Lys Ala Asn Ile Lys Gly Glu Arg Lys His
                665                 670                 675
Val Tyr Thr Asn Ala Lys Lys Gln Met Leu Ile Ser Leu Ala Pro
                680                 685                 690
Pro Val Leu Thr Leu His Leu Lys Arg Phe Gln Gln Ala Gly Phe
                695                 700                 705
Asn Leu Arg Lys Val Asn Lys His Ile Lys Phe Pro Glu Ile Leu
                710                 715                 720
Asp Leu Ala Pro Phe Cys Thr Leu Lys Cys Lys Asn Val Ala Glu
                725                 730                 735
Glu Asn Thr Arg Val Leu Tyr Ser Leu Tyr Gly Val Val Glu His
                740                 745                 750
Ser Gly Thr Met Arg Ser Gly His Tyr Thr Ala Tyr Ala Lys Ala
                755                 760                 765
Arg Thr Ala Asn Ser His Leu Ser Asn Leu Val Leu His Gly Asp
                770                 775                 780
Ile Pro Gln Asp Phe Glu Met Glu Ser Lys Gly Gln Trp Phe His
```

785 790 795
Ile Ser Asp Thr His Val Gln Ala Val Pro Thr Thr Lys Val Leu
800 805 810
Asn Ser Gln Ala Tyr Leu Leu Phe Tyr Glu Arg Ile Leu
815 820

<210> 13
<211> 404
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1920734CD1

<400> 13
Met Val Gln Leu Ala Pro Ala Ala Ala Met Asp Glu Val Thr Phe
1 5 10 15
Arg Ser Asp Thr Val Leu Ser Asp Val His Leu Tyr Thr Pro Asn
20 25 30
His Arg His Leu Met Val Arg Leu Asn Ser Val Gly Gln Pro Val
35 40 45
Phe Leu Ser Gln Phe Lys Leu Leu Trp Ser Gln Asp Ser Trp Thr
50 55 60
Asp Ser Gly Ala Lys Gly Gly Ser His Arg Asp Val His Thr Lys
65 70 75
Glu Pro Pro Ser Ala Glu Thr Gly Ser Thr Gly Ser Pro Pro Gly
80 85 90
Ser Gly His Gly Asn Glu Gly Phe Ser Leu Gln Ala Gly Thr Asp
95 100 105
Thr Thr Gly Gln Glu Val Ala Glu Ala Gln Leu Asp Glu Asp Gly
110 115 120
Asp Leu Asp Val Val Arg Arg Pro Arg Ala Ala Ser Asp Ser Asn
125 130 135
Pro Ala Gly Pro Leu Arg Asp Lys Val His Pro Met Ile Leu Ala
140 145 150
Gln Glu Glu Asp Asp Val Leu Gly Glu Glu Ala Gln Gly Ser Pro
155 160 165
His Asp Ile Ile Arg Ile Glu His Thr Met Ala Thr Pro Leu Glu
170 175 180
Asp Val Gly Lys Gln Val Trp Arg Gly Ala Leu Leu Leu Ala Asp
185 190 195
Tyr Ile Leu Phe Arg Gln Asp Leu Phe Arg Gly Cys Thr Ala Leu
200 205 210
Glu Leu Gly Ala Gly Thr Gly Leu Ala Ser Ile Ile Ala Ala Thr
215 220 225
Met Ala Arg Thr Val Tyr Cys Thr Asp Val Gly Ala Asp Leu Leu
230 235 240
Ser Met Cys Gln Arg Asn Ile Ala Leu Asn Ser His Leu Ala Ala
245 250 255
Thr Gly Gly Gly Ile Val Arg Val Lys Glu Leu Asp Trp Leu Lys
260 265 270
Asp Asp Leu Cys Thr Asp Pro Lys Val Pro Phe Ser Trp Ser Gln
275 280 285

```
Glu Glu Ile Ser Asp Leu Tyr Asp His Thr Thr Ile Leu Phe Ala
                290                 295                 300
Ala Glu Val Phe Tyr Asp Asp Asp Leu Thr Asp Ala Val Phe Lys
                305                 310                 315
Thr Leu Ser Arg Leu Ala His Arg Leu Lys Asn Ala Cys Thr Ala
                320                 325                 330
Ile Leu Ser Val Glu Lys Arg Leu Asn Phe Thr Leu Arg His Leu
                335                 340                 345
Asp Val Thr Cys Glu Ala Tyr Asp His Phe Arg Ser Cys Leu His
                350                 355                 360
Ala Leu Glu Gln Leu Thr Asp Gly Lys Leu Arg Phe Val Val Glu
                365                 370                 375
Pro Val Glu Ala Ser Phe Pro Gln Leu Leu Val Tyr Glu Arg Leu
                380                 385                 390
Gln Gln Leu Glu Leu Trp Lys Ile Ile Ala Glu Pro Val Thr
                395                 400
```

<210> 14
<211> 703
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2396858CD1

<400> 14

```
Met Ala Ala Ala Thr Gly Asp Pro Gly Leu Ser Lys Leu Gln Phe
  1               5                  10                  15
Ala Pro Phe Ser Ser Ala Leu Asp Val Gly Phe Trp His Glu Leu
                20                  25                  30
Thr Gln Lys Lys Leu Asn Glu Tyr Arg Leu Asp Glu Ala Pro Lys
                35                  40                  45
Asp Ile Lys Gly Tyr Tyr Tyr Asn Gly Asp Ser Ala Gly Leu Pro
                50                  55                  60
Ala Arg Leu Thr Leu Glu Phe Ser Ala Phe Asp Met Ser Ala Pro
                65                  70                  75
Thr Pro Ala Arg Cys Cys Pro Ala Ile Gly Thr Leu Tyr Asn Thr
                80                  85                  90
Asn Thr Leu Glu Ser Phe Lys Thr Ala Asp Lys Lys Leu Leu Leu
                95                 100                 105
Glu Gln Ala Ala Asn Glu Ile Trp Glu Ser Ile Lys Ser Gly Thr
               110                 115                 120
Ala Leu Glu Asn Pro Val Leu Leu Asn Lys Phe Leu Leu Leu Thr
               125                 130                 135
Phe Ala Asp Leu Lys Lys Tyr His Phe Tyr Tyr Trp Phe Cys Tyr
               140                 145                 150
Pro Ala Leu Cys Leu Pro Glu Ser Leu Pro Leu Ile Gln Gly Pro
               155                 160                 165
Val Gly Leu Asp Gln Arg Phe Ser Leu Lys Gln Ile Glu Ala Leu
               170                 175                 180
Glu Cys Ala Tyr Asp Asn Leu Cys Gln Thr Glu Gly Val Thr Ala
               185                 190                 195
Leu Pro Tyr Phe Leu Ile Lys Tyr Asp Glu Asn Met Val Leu Val
```

```
                200                 205                 210
Ser Leu Leu Lys His Tyr Ser Asp Phe Phe Gln Gly Gln Arg Thr
                215                 220                 225
Lys Ile Thr Ile Gly Val Tyr Asp Pro Cys Asn Leu Ala Gln Tyr
                230                 235                 240
Pro Gly Trp Pro Leu Arg Asn Phe Leu Val Leu Ala Ala His Arg
                245                 250                 255
Trp Ser Ser Ser Phe Gln Ser Val Glu Val Val Cys Phe Arg Asp
                260                 265                 270
Arg Thr Met Gln Gly Ala Arg Asp Val Ala His Ser Ile Ile Phe
                275                 280                 285
Glu Val Lys Leu Pro Glu Met Ala Phe Ser Pro Asp Cys Pro Lys
                290                 295                 300
Ala Val Gly Trp Glu Lys Asn Gln Lys Gly Gly Met Gly Pro Arg
                305                 310                 315
Met Val Asn Leu Ser Glu Cys Met Asp Pro Lys Arg Leu Ala Glu
                320                 325                 330
Ser Ser Val Asp Leu Asn Leu Lys Leu Met Cys Trp Arg Leu Val
                335                 340                 345
Pro Thr Leu Asp Leu Asp Lys Val Val Ser Val Lys Cys Leu Leu
                350                 355                 360
Leu Gly Ala Gly Thr Leu Gly Cys Asn Val Ala Arg Thr Leu Met
                365                 370                 375
Gly Trp Gly Val Arg His Ile Thr Phe Val Asp Asn Ala Lys Ile
                380                 385                 390
Ser Tyr Ser Asn Pro Val Arg Gln Pro Leu Tyr Glu Phe Glu Asp
                395                 400                 405
Cys Leu Gly Gly Gly Lys Pro Lys Ala Leu Ala Ala Ala Asp Arg
                410                 415                 420
Leu Gln Lys Ile Phe Pro Gly Val Asn Ala Arg Gly Phe Asn Met
                425                 430                 435
Ser Ile Pro Met Pro Gly His Pro Val Asn Phe Ser Ser Val Thr
                440                 445                 450
Leu Glu Gln Ala Arg Arg Asp Val Glu Gln Leu Glu Gln Leu Ile
                455                 460                 465
Glu Ser His Asp Val Val Phe Leu Leu Met Asp Thr Arg Glu Ser
                470                 475                 480
Arg Trp Leu Pro Ala Val Ile Ala Ala Ser Lys Arg Lys Leu Val
                485                 490                 495
Ile Asn Ala Ala Leu Gly Phe Asp Thr Phe Val Val Met Arg His
                500                 505                 510
Gly Leu Lys Lys Pro Lys Gln Gln Gly Ala Gly Asp Leu Cys Pro
                515                 520                 525
Asn His Pro Val Ala Ser Ala Asp Leu Leu Gly Ser Ser Leu Phe
                530                 535                 540
Ala Asn Ile Pro Gly Tyr Lys Leu Gly Cys Tyr Phe Cys Asn Asp
                545                 550                 555
Val Val Ala Pro Gly Asp Ser Thr Arg Asp Arg Thr Leu Asp Gln
                560                 565                 570
Gln Cys Thr Val Ser Arg Pro Gly Leu Ala Val Ile Ala Gly Ala
                575                 580                 585
Leu Ala Val Glu Leu Met Val Ser Val Leu Gln His Pro Glu Gly
                590                 595                 600
Gly Tyr Ala Ile Ala Ser Ser Ser Asp Asp Arg Met Asn Glu Pro
                605                 610                 615
```

```
Pro Thr Ser Leu Gly Leu Val Pro His Gln Ile Arg Gly Phe Leu
                620                 625                 630
Ser Arg Phe Asp Asn Val Leu Pro Val Ser Leu Ala Phe Asp Lys
                635                 640                 645
Cys Thr Ala Cys Ser Ser Lys Val Leu Asp Gln Tyr Glu Arg Glu
                650                 655                 660
Gly Phe Asn Phe Leu Ala Lys Val Phe Asn Ser Ser His Ser Phe
                665                 670                 675
Leu Glu Asp Leu Thr Gly Leu Thr Leu Leu His Gln Glu Thr Gln
                680                 685                 690
Ala Ala Glu Ile Trp Asp Met Ser Asp Asp Glu Thr Ile
                695                 700
```

<210> 15
<211> 145
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2634725CD1

<400> 15

```
Met Thr Leu Pro Ser Lys Gln Pro Gly Ser Gln Pro Arg Pro Ala
  1               5                  10                  15
Leu Ser Pro Gly Thr Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile
                20                  25                  30
Arg Val Ile Asn Gln Thr Thr Cys Glu Asn Leu Leu Pro Gln Gln
                35                  40                  45
Ile Thr Pro Arg Met Met Cys Val Gly Phe Leu Ser Gly Gly Val
                50                  55                  60
Asp Ser Cys Gln Val Ala Pro Gly Ala Gly Gly Arg Gln Val Gly
                65                  70                  75
Pro Gly Arg Gly Gly Thr Gly Asp Ser Pro Ala Gly Leu Val Ser
                80                  85                  90
Ala Gln Gly Asp Ser Gly Gly Pro Leu Ser Ser Val Glu Ala Asp
                95                 100                 105
Gly Arg Ile Phe Gln Ala Gly Val Val Ser Trp Gly Asp Gly Cys
               110                 115                 120
Ala Gln Arg Asn Lys Pro Gly Val Tyr Thr Arg Leu Pro Leu Phe
               125                 130                 135
Arg Asp Trp Ile Lys Glu Asn Thr Gly Val
               140                 145
```

<210> 16
<211> 518
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2643110CD1

<400> 16

Met Arg Lys Val Lys Lys Leu Arg Leu Asp Lys Glu Asn Thr Gly
1 5 10 15
Ser Trp Arg Ser Phe Ser Leu Asn Ser Glu Gly Ala Glu Arg Met
20 25 30
Ala Thr Thr Gly Thr Pro Thr Ala Asp Arg Cys Asp Ala Ala Ala
35 40 45
Thr Asp Asp Pro Ala Ala Arg Phe Gln Val Gln Lys His Ser Trp
50 55 60
Asp Gly Leu Arg Ser Ile Ile His Gly Ser Arg Lys Tyr Ser Gly
65 70 75
Leu Ile Val Asn Lys Ala Pro His Asp Phe Gln Phe Val Gln Lys
80 85 90
Thr Asp Glu Ser Gly Pro His Ser His Arg Leu Tyr Tyr Leu Gly
95 100 105
Met Pro Tyr Gly Ser Arg Glu Asn Ser Leu Leu Tyr Ser Glu Ile
110 115 120
Pro Lys Lys Val Arg Lys Glu Ala Leu Leu Leu Leu Ser Trp Lys
125 130 135
Gln Met Leu Asp His Phe Gln Ala Thr Pro His His Gly Val Tyr
140 145 150
Ser Arg Glu Glu Glu Leu Leu Arg Glu Arg Lys Arg Leu Gly Val
155 160 165
Phe Gly Ile Thr Ser Tyr Asp Phe His Ser Glu Ser Gly Leu Phe
170 175 180
Leu Phe Gln Ala Ser Asn Ser Leu Phe His Cys Arg Asp Gly Gly
185 190 195
Lys Asn Gly Phe Met Val Ser Pro Met Lys Pro Leu Glu Ile Lys
200 205 210
Thr Gln Cys Ser Gly Pro Arg Met Asp Pro Lys Ile Cys Pro Ala
215 220 225
Asp Pro Asp Phe Phe Ser Phe Ile Asn Asn Ser Asp Leu Trp Val
230 235 240
Ala Asn Ile Glu Thr Gly Glu Glu Arg Arg Leu Thr Phe Cys His
245 250 255
Gln Gly Leu Ser Asn Val Leu Asp Asp Pro Lys Ser Ala Gly Val
260 265 270
Ala Thr Phe Val Ile Gln Glu Glu Phe Asp Arg Phe Thr Gly Tyr
275 280 285
Trp Trp Cys Pro Thr Ala Ser Trp Glu Gly Ser Glu Gly Leu Lys
290 295 300
Thr Leu Arg Ile Leu Tyr Glu Glu Val Asp Glu Ser Glu Val Glu
305 310 315
Val Ile His Val Pro Ser Pro Ala Leu Glu Glu Arg Lys Thr Asp
320 325 330
Ser Tyr Arg Tyr Pro Arg Thr Gly Ser Lys Asn Pro Lys Ile Ala
335 340 345
Leu Lys Leu Ala Glu Phe Gln Thr Asp Ser Gln Gly Lys Ile Val
350 355 360
Ser Thr Gln Glu Lys Glu Leu Val Gln Pro Phe Ser Ser Leu Phe
365 370 375
Pro Lys Val Glu Tyr Ile Ala Arg Ala Gly Trp Thr Arg Asp Gly
380 385 390
Lys Tyr Ala Trp Ala Met Phe Leu Asp Arg Pro Gln Gln Trp Leu
395 400 405

```
Gln Leu Val Leu Leu Pro Pro Ala Leu Phe Ile Pro Ser Thr Glu
                410                 415                 420
Asn Glu Glu Gln Arg Leu Ala Ser Ala Arg Ala Val Pro Arg Asn
                425                 430                 435
Val Gln Pro Tyr Val Val Tyr Glu Glu Val Thr Asn Val Trp Ile
                440                 445                 450
Asn Val His Asp Ile Phe Tyr Pro Phe Pro Gln Ser Glu Gly Glu
                455                 460                 465
Asp Glu Leu Cys Phe Leu Arg Ala Asn Glu Cys Lys Thr Gly Phe
                470                 475                 480
Cys His Leu Tyr Lys Val Thr Ala Val Leu Lys Ser Gln Gly Tyr
                485                 490                 495
Asp Trp Ser Glu Pro Phe Ser Pro Gly Glu Gly Glu Gln Ser Leu
                500                 505                 510
Thr Asn Ala Val Asp Ser Ser Arg
                515
```

<210> 17
<211> 476
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2701396CD1

<400> 17

```
Met Trp Thr Gly Tyr Lys Ile Leu Ile Phe Ser Tyr Leu Thr Thr
  1               5                  10                  15
Glu Ile Trp Met Glu Lys Gln Tyr Leu Ser Gln Arg Glu Val Asp
                20                  25                  30
Leu Glu Ala Tyr Phe Thr Arg Asn His Thr Val Leu Gln Gly Thr
                35                  40                  45
Arg Phe Lys Arg Ala Ile Phe Gln Gly Gln Tyr Cys Arg Asn Phe
                50                  55                  60
Gly Cys Cys Glu Asp Arg Asp Asp Gly Cys Val Thr Glu Phe Tyr
                65                  70                  75
Ala Ala Asn Ala Leu Cys Tyr Cys Asp Lys Phe Cys Asp Arg Glu
                80                  85                  90
Asn Ser Asp Cys Cys Pro Asp Tyr Lys Ser Phe Cys Arg Glu Glu
                95                 100                 105
Lys Glu Trp Pro Pro His Thr Gln Pro Trp Tyr Pro Glu Gly Cys
               110                 115                 120
Phe Lys Asp Gly Gln His Tyr Glu Glu Gly Ser Val Ile Lys Glu
               125                 130                 135
Asn Cys Asn Ser Cys Thr Cys Ser Gly Gln Gln Trp Lys Cys Ser
               140                 145                 150
Gln His Val Cys Leu Val Arg Ser Glu Leu Ile Glu Gln Val Asn
               155                 160                 165
Lys Gly Asp Tyr Gly Trp Thr Ala Gln Asn Tyr Ser Gln Phe Trp
               170                 175                 180
Gly Met Thr Leu Glu Asp Gly Phe Lys Phe Arg Leu Gly Thr Leu
               185                 190                 195
Pro Pro Ser Pro Met Leu Leu Ser Met Asn Glu Met Thr Ala Ser
```

200 205 210
Leu Pro Ala Thr Thr Asp Leu Pro Glu Phe Leu Leu Leu Leu Ile
215 220 225
Asn Gly Leu Asp Gly Leu Met Ala His Trp Ile Lys Lys Ile Cys
230 235 240
Ala Ala Ser Trp Ala Phe Ser Thr Ala Ser Val Ala Ala Asp Arg
245 250 255
Ile Ala Ile Gln Ser Lys Gly Arg Tyr Thr Ala Asn Leu Ser Pro
260 265 270
Gln Asn Leu Ile Ser Cys Cys Ala Lys Asn Arg His Gly Cys Asn
275 280 285
Ser Gly Ser Ile Asp Arg Ala Trp Trp Tyr Leu Arg Lys Arg Gly
290 295 300
Leu Val Ser His Ala Cys Tyr Pro Leu Phe Lys Asp Gln Asn Ala
305 310 315
Thr Asn Asn Gly Cys Ala Met Ala Ser Arg Ser Asp Gly Arg Gly
320 325 330
Lys Arg His Ala Thr Lys Pro Cys Pro Asn Asn Val Glu Lys Ser
335 340 345
Asn Arg Ile Tyr Gln Cys Ser Pro Pro Tyr Arg Val Ser Ser Asn
350 355 360
Glu Thr Glu Ile Met Lys Glu Ile Met Gln Asn Gly Pro Val Gln
365 370 375
Ala Ile Met Gln Val Arg Glu Asp Phe Phe His Tyr Lys Thr Gly
380 385 390
Ile Tyr Arg His Val Thr Ser Thr Asn Lys Glu Ser Glu Lys Tyr
395 400 405
Arg Lys Leu Gln Thr His Ala Val Lys Leu Thr Gly Trp Gly Thr
410 415 420
Leu Arg Gly Ala Gln Gly Gln Lys Glu Lys Phe Trp Ile Ala Ala
425 430 435
Asn Ser Trp Gly Lys Ser Trp Gly Glu Asn Gly Tyr Phe Arg Ile
440 445 450
Leu Arg Gly Val Asn Glu Ser Asp Ile Glu Lys Leu Ile Ile Ala
455 460 465
Ala Trp Gly Gln Leu Thr Ser Ser Asp Glu Pro
470 475

<210> 18
<211> 229
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3134404CD1

<400> 18
Met Pro Cys Ala Gln Arg Ser Trp Leu Ala Asn Leu Ser Val Val
1 5 10 15
Ala Gln Leu Leu Asn Phe Gly Ala Leu Cys Tyr Gly Arg Gln Leu
20 25 30
Gln Pro Gly Pro Val Arg Phe Pro Asp Arg Arg Gln Glu His Phe
35 40 45

```
Ile Lys Gly Leu Pro Glu Tyr His Val Val Gly Pro Val Arg Val
                50                  55                  60
Asp Ala Ser Gly His Phe Leu Ser Tyr Gly Leu His Tyr Pro Ile
                65                  70                  75
Thr Ser Ser Arg Arg Lys Arg Asp Leu Asp Gly Ser Glu Asp Trp
                80                  85                  90
Val Tyr Tyr Arg Ile Ser His Glu Glu Lys Asp Leu Phe Phe Asn
                95                 100                 105
Leu Thr Val Asn Gln Gly Phe Leu Ser Asn Ser Tyr Ile Met Glu
               110                 115                 120
Lys Arg Tyr Gly Asn Leu Ser His Val Lys Met Met Ala Ser Ser
               125                 130                 135
Ala Pro Leu Cys His Leu Ser Gly Thr Val Leu Gln Gln Gly Thr
               140                 145                 150
Arg Val Gly Thr Ala Ala Leu Ser Ala Cys His Gly Leu Thr Gly
               155                 160                 165
Phe Phe Gln Leu Pro His Gly Asp Phe Phe Ile Glu Pro Val Lys
               170                 175                 180
Lys His Pro Leu Val Glu Gly Gly Tyr His Pro His Ile Val Tyr
               185                 190                 195
Arg Arg Gln Lys Val Pro Glu Thr Lys Glu Pro Thr Cys Gly Leu
               200                 205                 210
Lys Gly Ile Val Thr His Met Ser Ser Trp Val Glu Glu Ser Val
               215                 220                 225
Leu Phe Phe Trp
```

<210> 19
<211> 3159
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 155179CB1

<400> 19

```
gcagggacga cgcctgtgag acccgcgagc ggcctcgggg accatgggga gcgatcgggc  60
ccgcaagggg agggggccga agacttcggc gcgggactca agtacaactc ccggcacgag 120
aaagtgaatg gcttggagga aggcgtggag ttcctgccag tcaacaacgt caagaaggtg 180
gaaaagcatg gcccggggcg ctgggtggtg ctggcagccg tgctgatcgg cctcctcttg 240
gtcttgctgg ggatcggctt cctggtgtgg catttgcagt accgggacgt gcgtgtccag 300
aaggtcttca atggctacat gaggatcaca aatgagaatt ttgtggatgc ctacgagaac 360
tccaactcca ctgagtttgt aagcctggcc agcaaggtga aggacgcgct gaagctgctg 420
tacagcggag tcccattcct gggcccctgc cacaaggagt cggctgtgac ggccttcagc 480
gagggcagcg tcatcgccta ctactggtct gagttcagca tcccgcagca cctggtggag 540
gaggccgagc gcgtcatggc cgaggagcgc gtagtcatgc tgcccccgcg ggcgcgctcc 600
ctgaagtcct ttgtggtcac ctcagtggtg gctttcccca cggactccaa aacagtacag 660
aggacccagg acaacagctg cagctttggc ctgcacgccc gcggtgtgga gctgatgcgc 720
ttcaccacgc ccggcttccc tgacagcccc taccccgctc atgcccgctg ccagtgggcc 780
ctgcgggggg acgccgactc agtgctgagc ctcaccttcc gcagctttga ccttgcgtcc 840
tgcgacgagc gcggcagcga cctggtgacg gtgtacaaca ccctgagccc catggagccc 900
cacgccctgg tgcagttgtg tggcacctac cctccctcct acaacctgac cttccactcc 960
tcccagaacg tcctgctcat cacactgata accaacactg agcggcggca tcccggcttt 1020
gaggccacct tcttccagct gcctaggatg agcagctgtg gaggccgctt acgtaaagcc 1080
```

caggggacat tcaacagccc ctactaccca ggccactacc cacccaacat tgactgcaca 1140
tggaacattg aggtgcccaa caaccagcat gtgaaggtgc gcttcaaatt cttctacctg 1200
ctggagcccg gcgtgcctgc gggcacctgc cccaaggact acgtggagat caacggggag 1260
aaatactgcg gagagaggtc ccagttcgtc gtcaccagca acagcaacaa gatcacagtt 1320
cgcttccact cagatcagtc ctacaccgac accggcttct tagctgaata cctctcctac 1380
gactccagtg acccatgccc ggggcagttc acgtgccgca cggggcggtg tatccggaag 1440
gagctgcgct gtgatggctg ggccgactgc accgaccaca gcgatgagct caactgcagt 1500
tgcgacgccg gccaccagtt cacgtgcaag aacaagttct gcaagcccct cttctgggtc 1560
tgcgacagtg tgaacgactg cggagacaac agcgacgagc aggggtgcag ttgtccggcc 1620
cagaccttca ggtgttccaa tgggaagtgc ctctcgaaaa gccagcagtg caatgggaag 1680
gacgactgtg gggacgggtc cgacgaggcc tcctgcccca aggtgaacgt cgtcacttgt 1740
accaaacaca cctaccgctg cctcaatggg ctctgcttga gcaagggcaa ccctgagtgt 1800
gacgggaagg aggactgtag cgacggctca gatgagaagg actgcgactg tgggctgcgg 1860
tcattcacga gacaggctcg tgttgttggg ggcacggatg cggatgaggg cgagtggccc 1920
tggcaggtaa gcctgcatgc tctgggccag ggccacatct gcggtgcttc cctcatctct 1980
cccaactggc tggtctctgc cgcacactgc tacatcgatg acagaggatt caggtactca 2040
gaccccacgc agtggacggc cttcctgggc ttgcacgacc agagccagcg cagcgccct 2100
ggggtgcagg agcgcaggct caagcgcatc atctcccacc ccttcttcaa tgacttcacc 2160
ttcgactatg acatcgcgct gctggagctg gagaaaccgg cagagtacag ctccatggtg 2220
cggcccatct gcctgccgga cgcctcccat gtcttccctg ccggcaaggc catctgggtc 2280
acgggctggg gacacaccca gtatggaggc actggcgcgc tgatcctgca aaagggtgag 2340
atccgcgtca tcaaccagac cacctgcgag aacctcctgc cgcagcagat cacgccgcgc 2400
atgatgtgcg tgggcttcct cagcggcggc gtggactcct gccagggtga ttccgggga 2460
cccctgtcca gcgtggaggc ggatgggcgg atcttccagg ccggtgtggt gagctgggga 2520
gacggctgcg ctcagaggaa caagccaggc gtgtacacaa ggctccctct gtttcgggac 2580
tggatcaaag agaacactgg ggtatagggg ccggggccac ccaaatgtgt acacctgcgg 2640
ggccacccat cgtccacccc agtgtgcacg cctgcaggct ggagactgga ccgctgactg 2700
caccagcgcc cccagaacat acactgtgaa ctcaatctcc agggctccaa atctgcctag 2760
aaaacctctc gcttcctcag cctccaaagt ggagctggga ggtagaaggg gaggacactg 2820
gtggttctac tgacccaact gggggcaaag gtttgaagac acagcctccc ccgccagccc 2880
caagctgggc cgaggcgcgt ttgtgtatat ctgcctcccc tgtctgtaag gagcagcggg 2940
aacggagctt cggagcctcc tcagtgaagg tggtggggct gccggatctg ggctgtgggg 3000
cccttgggcc acgctcttga ggaagcccag gctcggagga ccctggaaaa cagacgggtc 3060
tgagactgaa attgttttac cagctcccag ggtggacttc agtgtgtgta tttgtgtaaa 3120
tgagtaaaac attttatttc tttttaaaaa aaaaaaaaa 3159

<210> 20
<211> 1355
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2415780CB1

<400> 20
tgaatttaat cctactcact atagggaatt tggccctcga ggccaagaat tcggcacgag 60
gtggaaagct gcggcccagc gcggactagt gaggacctcc acagctcctg acattgccag 120
gagtcctgtc ggcgttttct cccagcctcc gccatgccgg cggtgctggg ttttgaaggc 180
agcgccaata agattggcgt gggcgtggtg cgggatggca aggtgctggc gaacccgcgg 240
cggacttacg tcacgcctcc tggcacagga ttccttccag gtgatacagc caggcatcac 300
cgagctgtta tcctagacct gctgcaggag gcactaacag agtctggatt aacctcccag 360
gatatcgact gcattgcata caccaagggc cctggcatgg gtgccccact ggtttctgtg 420
gctgttgtgg cccgtactgt ggcccagctg tggaataagc cattggtggg tgtgaaccac 480

tgtataggcc acattgagat gggccgcctc atcactggag ccaccagccc aaccgtgttg 540
tatgtgagtg gaggaaatac gcaggtgatt gcatactcgg aacatcgtta ccgtatcttt 600
ggggaaacca tcgatattgc agtgggtaat tgtctggatc gttttgctcg agtgctgaag 660
atttctaacg acccaagtcc aggatacaac attgaacaga tggcaaagcg aggcaagaag 720
ctggttgagc tgccatacac tgtaaagggg atggacgtct cattctcagg gatcctgtct 780
ttcattgagg atgtagccca tcggatgctg gccacaggcg agtgtactcc tgaggatctg 840
tgtttctccc tgcaggaaac tgtgtttgca atgctggtag agatcacaga gcgagccatg 900
gcacattgtg gctcccagga ggccctcatt gtgggaggag tggggtgtaa tgtgaggcta 960
caggagatga tggcaacaat gtgccaggaa cgtggagccc ggcttttgc tacagatgag 1020
agattctgta ttgacaatgg agcgatgata gcccaggctg gctgggagat gtttcgggct 1080
ggacacagga ccccactcag tgattctggg gttacacaga ggtatcggac agatgaagta 1140
gaggtgacct ggagggacta ataagatcaa cagaatcaga gtagatagtt ccttaatcgg 1200
aacccaaagg accccgtgcc tcaatctcta tcctgatgtc atgggagtcc tagcaaagct 1260
atagactcca agcaaggctt ggggtccttt atggaacccc aggatgactc agcaataaaa 1320
tatttttggt tttttggttt tgtaaaaaaa aaaaa 1355

<210> 21
<211> 1601
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2879274CB1

<400> 21
ccccgcccac gtgacgggcg cccgcggaag gcgacatggg ctccgctccc tgggccccgg 60
tcctgctgct ggcgctcggg ctgcgcggcc tccaggcggg ggcccgcagg tccggatccc 120
cggcttccag gagcgcttct tccagcagcg tctggaccac ttcaacttcg agcgcttcgg 180
caacaagacc ttccctcagc gcttcctggt gtcgggttct gggtccgggg cgaggggccc 240
atcttcttct acactgggaa cgagggcgac gtgtgggcct tcgccaacaa ctcgggcttc 300
gtcgcggact ggcggccgag cgggggggctc tactggtctt cgcggacacc gctactacgg 360
gaagtcgctg ccgttcggtg cgcagtccac gcagcgcggg cacacggagc tgctgacggt 420
ggagcaggcc ctggccgact tcgcagagct gctccgcgcg ctacgacgcg acctcggggc 480
ccaggatgcc cccgccatcg ccttcggtgg aagttatggg gggatgctca gtgcctacct 540
gaggatgaag tatccccacc tggtggcggg ggcgctggcg gccagcgcgc ccgttctagc 600
tgtggcaggc ctcggcgact ccaaccagtt cttccgggac gtcacggcgg actttgaggg 660
ccagagtccc aaatgcaccc agggtgtgcg ggaagcgttc cgacagatca aggacttgtt 720
cctacaggga gcctacgaca cggtccgctg ggagttcggc acctgccagc cgctgtcaga 780
cgagaaggac ctgacccagc tcttcatgtt cgcccggaat gccttcaccg tgctggccat 840
gatggactac ccctacccca ctgacttcct gggtcccctc cctgccaacc ccgtcaaggt 900
gggctgtgat cggctgctga gtgaggccca gaggatcacg gggctgcgag cactggcagg 960
gctggtctac aacgcctcgg gctccgagca ctgctacgac atctaccggc tctaccacag 1020
ctgtgctgac cccactggct gcggcaccgg ccccgacgcc agggcctggg actaccaggc 1080
ctgcaccgag atcaacctga ccttcgccag caacaatgtg accgatatgt tccccgacct 1140
gcccttcact gacgagctcc gccagcggta ctgcctggac acctggggcg tgtggccccg 1200
gcccgactgg ctgctgacca gcttctgggg gggtgatctt agagccgcca gcaacatcat 1260
cttctccaac gggaacctgg acccctgggc aggggggcggg attcggagga acctgagtgc 1320
ctcagtcatc gccgtcacca tccaggggggg agcgcaccac ctcgacctca gagcctccca 1380
cccagaagat cctgcttccg tggttgaggc gcggaagctg gaggccacca tcatcggcga 1440
gtgggtaaag gcagccaggc gtgagcagca gccagctctg cgtgggggggc ccagactcag 1500
cctctgagca caggactgga ggggtctcaa ggctcctcat ggagtggggg cttcactcaa 1560
gcagctggcg gcagagggaa ggggctgaat aaacgcctgg t 1601

<210> 22
<211> 2364
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 358050CB1

<400> 22
tggcaagatg gcggccatgg agaccgagac ggcgccgctg accctagagt cgctgcccac 60
cgatcccctg ctcctcatct tatccttttt ggactatcgg gatctaatca actgttgtta 120
tgtcagtcga agacttagcc agctatcaag tcatgatccg ctgtggagaa gacattgcaa 180
aaaatactgg ctgatatctg aggaagagaa aacacagaag aatcagtgtt ggaaatctct 240
cttcatagat acttactctg atgtaggaag atacattgac cattatgctg ctattaaaaa 300
ggcctgggat gatctcaaga aatatttgga gcccaggtgt cctcggatgg ttttatctct 360
gaaagagggt gctcgagagg aagacctcga tgctgtggaa gcgcagattg gctgcaagct 420
tcctgacgat tatcgatgtt cataccgaat tcacaatgga cagaagttag tggttcctgg 480
gttattggga agcatggcac tgtctaatca ctatcgttct gaagatttgt tagacgtcga 540
tacagctgcc ggaggattcc agcagagaca gggactgaaa tactgtctcc ctttaacttt 600
ttgcatacat actggtttga gtcagtacat agcagtggaa gctgcagagg gccgaaacaa 660
aaatgaagtt ttctaccaat gtccagacca aatggctcga aatccagctg ctattgacat 720
gtttattata ggtgctactt ttactgactg gtttacctct tatgtcaaaa atgttgtatc 780
aggtggcttc cccatcatca gagaccaaat tttcagatat gttcacgatc cagaatgtgt 840
agcaacaact ggggatatta ctgtgtcagt ttccacatcg tttctgccag aacttagctc 900
tgtacatcca ccccactatt tcttcacata ccgaatcagg attgaaatgt caaaagatgc 960
acttcctgag aaggcctgtc agttggacag tcgctattgg agaataacaa atgctaaggg 1020
tgacgtggaa gaagttcaag gacctggagt agttggtgaa tttccaatca tcagcccagg 1080
tcgggtatat gaatacacaa gctgtaccac attctctaca acatcaggat acatggaagg 1140
atattatacc ttccattttc tttactttaa agacaagatc tttaatgttg ccattccccg 1200
attccatatg gcatgtccaa cattcagggt gtctatagcc cgattggaaa tgggtcctga 1260
tgaatatgaa gagatggaag aagaggagga ggaggaagag gaggaagacg aggatgatga 1320
ttcagcagat atggatgaat cagatgaaga tgatgaagag gagagacgga ggagagtctt 1380
tgatgttccc attcgcagac gccgctgctc acgccttttt tagcaagcct tctgctgatg 1440
gaagcactag gatgattcta ggctgttaaa tagatttctc aataatgtaa ataactaaat 1500
tgttctctgc atatagcagg aaaactagca tgaaatattg tttcaggccc tgggttctat 1560
gtgacactac attaggaatt ggattgtttg ggtttgcttt gtgtttttga ggtagaggaa 1620
gaaatgggaa tctttttttt ctcttccagg agtcagtgga agaatagttc tctagctaag 1680
gaacggacat acctttgttt taaaatattt tatacttaca aaaatctaga aatggagagg 1740
gaactgtttt gaataaggat ttaaaatacc tgcacaagga tagagagaaa ctatgtgact 1800
cattctgtga aaagacttct tgcagttgtg agttatttag aaatgatcaa aatttgtaat 1860
taggctaatc catttagtga ttcctaatat tttgtactca cagagaacta attgactaaa 1920
caacttgaac gctagtggtt tgtccttaga caatctgtct ttgaatttaa agtctttatc 1980
gctaagacct tgactttaaa ttttttcatca ctacaacctt gaatttaatt tcaggtcttc 2040
aacatgatga ccttggattt aatttaaagt cttcaacact atgcgcttta tcatattatt 2100
cacagatgca tttttgaaat gtagtatgta aaagtatgta acgtgctgtt tattaacaaa 2160
agattgttca caacatctca tgtagtttaa atttgtaaat actgcttctg ttttgtttct 2220
cctttataca cttgactgtc tttgtgataa gtgacatgaa ttttatgtta ggattaagta 2280
tgttttcctg aaacttggat ttttttttgta attatataat tgagagttaa gaatgaaatc 2340
cttcaagtgt taaaaactcg acat 2364

<210> 23
<211> 531
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 700745CB1

<400> 23
agcagcggca atgacccctt ggctcgggct catcgtgctc ctgggcagct ggagcctggg 60
ggactggggc gccgaggcgt gcacatgctc gcccagccac cccaggacg ccttctgcaa 120
ctccgacatc ggtaagcgct cctggtgccc cgcccgagcc ccacgctgca gccaggactg 180
cagcgctgct tagggaggca gggcgagccc cactcctttc ctctgcccca ggagaggggc 240
agacggggtt ggggcggagt ggagaaactc gatgtccttg ggcgggggcg ctggcatagc 300
tgagagggga agatgccctg cagaggaaac tcacagtggc tgagggagcc cctggccgcc 360
tttgctttcc taacttaggt cgtgaggttc ctaccggtcc ttttgacatc tggaaaatgt 420
ccccattcac tactaacgga ggaagggcta gaagagaagg gtggggaaag ggttcccaaa 480
acttggaatg ctaactaaag tgctgggaaa ttgaatagtt aaaaaaaaaa a 531

<210> 24
<211> 1181
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2026480CB1

<400> 24
gcccgcttga ggcgtagggg gtggcgctct ccgttcggcg gcgctcccat ggcgcacatt 60
accattaacc agtacctgca gcaggtgtac gaagccatcg acagcagaga tggagcatct 120
tgtgcagagt tggtgtcttt taaacatcct catgttgcaa acccacgact tcaaatggcc 180
tctccagagg agaagtgtca acaagtcttg gaacccctt atgatgaaat gtttgcagct 240
catttaaggt gcacttatgc agtggggaat catgacttca tagaggcata caagtgccag 300
accgtgatag tccaatcatt cttgcgagca ttccaggccc acaaagaaga aaactgggct 360
ctgcctgtca tgtatgcagt agcgcttgac cttcgagtgt ttgccaataa tgcagatcaa 420
cagttggtaa agaaaggaaa aagcaaagtt ggggacatgt tggaaaaagc agcagagtta 480
ctgatgagct gtttccgggt ctgtgccagc gacacccgtg ctggtataga ggactctaag 540
aagtggggca tgctgtttct ggtgaaccag ctgtttaaaa tctactttca tgcaggtgga 600
ggacgtggac attgacgaag ttcagtgtat tctggctaac ttgatataca tgggacacgt 660
caaaggctac atatcgcatc agcatcagaa gctggtggtc agcaagcaga acccatttcc 720
tcccctgtcc acggtgtgtt gaaagtacac ggagccccga ggacggcctg atgaaaacct 780
ggaattcctt tttcacagac tcggctggtt ctggagtctt tgtgagactt ctttgaagga 840
ggctttgcgt gaaggctgct cggctcactt ttcctaagtg tggttcctga aggctgtctt 900
tgtaactttt tgtagttctt tgtgtaaaaa gcgtattctg aatttataca catggcatgt 960
tcttcattat atcttccagg atacatctat ttttatatat taaatttgaa tgtgttatca 1020
aaatgcttgg ttaacttaag gcacctttt aaaagcagaa tttaatttga tttaaatttt 1080
ccagatttta tagcttgcct gtatggatgc tcctcaattt atgatggggt tacatcccaa 1140
taaacttatt ttatttgcct ttgaaaaaga aaaaaaaaaa a 1181

<210> 25
<211> 1617
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2132401CB1

<400> 25
tgaactgggg aacagatatc agcttccagt gtattgcttg gcaccgccaa tcaacatgat 60
agaggaaaag agcgacatag agactctgga tattcctgag ccaccaccca attctggata 120
tgaatgtcag cttcgtttgc gcctttccac aggcaaagac ctcaagcttg tggttcgcag 180
cacagacaca gtattccaca tgaagagacg gttgcatgca gcagagggag tggaaccagg 240
tagtcagcgg tggtttttt ctggcagacc tctcactgac aaaatgaagt tcgaagagct 300
gaagatccca aaggactatg ttgtacaggt tatagtgagc caacctgtgc agaacccaac 360
accagtggag aactgaactg agccctgttg gccagctccc acatccctct gctccttttt 420
atggttcttg ttgtcatttc ctactctgcg gcgtgaaatc tatttcactg ctctaaattc 480
cctatgaatg gatttagttc tgaggaatta ccagtgaaaa attccatctg tgatggagac 540
caacaaaaat aataaaacac aaagagccag gctttgagac tcatgtaatt acaatttcta 600
atttgaaagg cagttaagaa atagataacc atttatttta gaacactcaa caactatgta 660
atggctatat ttcagtgact tggactgtaa atgaaacatt gcatccatga aggacagcac 720
caagcacctt tttgaataca gaatttttt agaaaaatat atcaaattat ataatttcca 780
gaaaccataa ataatggata taaaacttaa ccttttttgtt ttgttttgtt ttgttttgtt 840
ttgttttggt taatggaaac tgaaaagagc agtatttgag gttgcttcta ttctggtttt 900
ttattcttag ctcaattaat atttagccat aaatgagtag agactggcaa attgtgcttt 960
agtgttgctt tcctcatccc cacatcttga gctccttatt tacattcaca ctaaattttg 1020
gtgccttcca gcacattagt ggcaggcacc cttctggaac actaggcaat aatttcatca 1080
atacagtcag gtctcttgag tttcaacaga tactcagttg aaaagtcgct gtcatcttgc 1140
tgcataagta ttttgaaagg tctgtataac gaagccattt ttatatccag gcttagaagg 1200
tcactactat atagtacctt cattgatcta tctattctgc ttggaacttt tcatagctaa 1260
gtataacccc caaatgcatg gttcctgggt caggagacac caaaatcaat tatagctgtt 1320
cccactcaag ttaataaagt aaatgatttc ctccactttg catggagggg tgtaaggaaa 1380
gcctatttta tctgtgcctg ggagaactgt gcctattttc agtcttttag aggaaatttc 1440
aactcaaaat ttttaagtgt gaaaggttta ctggtgtcac ataaacatta gcagtgagac 1500
caaataatga aacattgctt tataccttag tgctttccag ttcactgtta cctctagtca 1560
tggtagatga caattttcct ccctcatctt ttgtagcaaa gaagcaaatt aagaggg 1617

<210> 26
<211> 1661
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2568875CB1

<400> 26
ggacaccagt gatgctcctg ggaccctacg caatctgcgc ctgcgtctca tcagtcgccc 60
cacatgtaac tgtatctaca accagctgca ccagcgacac ctgtccaacc cggcccggcc 120
tgggatgcta tgtgggggcc cccagcctgg ggtgcagggc ccctgtcagg tctgataggg 180
agaagagaag gagcagaagg ggaggggcct aaccctgggc tgggggttgg actcacagga 240
ctgggggaaa gagctgcaat cagagggtgt ctgccatagc tgggctcagg catctgtcct 300
tggctttgtt gcctggctcc agggagattc cgggggccct gtgctgtgcc tcgagcctga 360
cggacactgg gttcaggctg gcatcatcag ctttgcatca agctgtgccc aggaggacgc 420
tcctgtgctg ctgaccaaca cagctgctca cagttcctgg ctgcaggctc gagttcaggg 480
ggcagctttc ctggcccaga gcccagagac cccggagatg agtgatgagg acagctgtgt 540
agcctgtgga tccttgagga cagcaggtcc ccaggcagga gcaccctccc catggccctg 600
ggaggccagg ctgatgcacc agggacagct ggcctgtggc ggagccctgg tgtcagagga 660

ggcggtgcta actgctgccc actgcttcat tgggcgccag gccccagagg aatggagcgt 720
agggctgggg accagaccgg aggagtgggg cctgaagcag ctcatcctgc atggagccta 780
cacccaccct gaggggggct acgacatggc cctcctgctg ctggcccagc ctgtgacact 840
gggagccagc ctgcggcccc tctgcctgcc ctatgctgac caccacctgc ctgatgggga 900
gcgtggctgg gttctgggac gggcccgccc aggagcaggc atcagctccc tccagacagt 960
gcccgtgacc ctcctggggc ctagggcctg cagccggctg catgcagctc ctgggggtga 1020
tggcagccct attctgccgg ggatggtgtg taccagtgct gtgggtgagc tgcccagctg 1080
tgagggcctg tctggggcac cactggtgca tgaggtgagg ggcacatggt tcctggccgg 1140
gctgcacagc ttcggagatg cttgccaagg ccccgccagg ccggcggtct tcaccgcgct 1200
ccctgcctat gaggactggg tcagcagttt ggactggcag gtctacttcg ccgaggaacc 1260
agagcccgag gctgagcctg gaagctgcct ggccaacata agccaaccaa ccagctgctg 1320
acaggggacc tggccattct caggacaaga gaatgcaggc aggcaaatgg cattactgcc 1380
cctgtcctcc ccaccctgtc atgtgtgatt ccaggcacca gggcaggccc agaagcccag 1440
cagctgtggg aaggaacctg cctggggcca caggtgccca ctccccaccc tgcaggacag 1500
gggtgtctgt ggacactccc acacccaact ctgctaccaa gcaggcgtct cagctttcct 1560
cctcctttac cctttcagat acaatcacgc cagccacgtt gttttgaaaa tttctttttt 1620
tgggggggcag cagttttcct ttttttaaac ttaaataaat t 1661

<210> 27
<211> 1010
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3408908CB1

<400> 27
ctttcctctc ctgactaagt ttctctggct tccctgaggc tgcaggtgtt aatctggggg 60
gccctgggcc ctgagccggc agcagaaata tgaggaccca gagccttctc ctcctggggg 120
ccctcctggc tgtggggagt cagctgcctg ctgtctttgg caggaagaag ggagagaaat 180
cgggggggctg cccgccagat gatgggccct gcctcctatc ggtgcctgac cagtgcgtgg 240
aagacagcca gtgtcccttg accaggaagt gctgctacag agcttgcttc cgccagtgtg 300
tccccagggt ctctgtgaag ctgggcagct gcccagagga ccaactgcgc tgcctcagcc 360
ccatgaacca cctgtgttac aaggactcag actgctcggg caaaaagcga tgctgccaca 420
gcgcctgcgg gcgggattgc cgggatcctg ccagaggcta attctgattt aggatctgtg 480
gctctgcacc taagctgggg accaacggaa agagttcacg atgggaggcc tggggccctg 540
cccgctggac agcactatct ctaccagcgg tggttccagc cttctgataa tcactggcct 600
gctgacactt ccctgcaacc catccacccc tggtttctcc tcctgggagt caaagtccat 660
agcctgagct cggaggaagg cctctgtatc accccagtac tctgcaccac tgccatacga 720
gcttcccacc cttcctaacg ctttcacacc aatccgtaca tgctgcttcc tccaccaaaa 780
atgcccaatt caggcagacc ctgacctctc cctcaggcag cccaaccatc cagaatgaat 840
attcttgcag agttttccaa acatcagtca ttcacctctt tcatgatttt caccatacct 900
acaaaatagc accatgatag gttgcacgct gcctgtacca ccatttactt aatgttttct 960
ttaaatggct cacttttgta tataaataaa ttcatttcaa aaaaaaaaaa 1010

<210> 28
<211> 1627
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 3772696CB1

<400> 28
gcttcagctg aagaaagaga ggaatgaagc gccttctgct tctgtgtttg ttctttataa 60
cattttcttc tgcatttccc ttagtccgga tgacggaaaa tgaagaaaat atgcaactgg 120
ctcaggcata tctcaaccag ttctactctc ttgaaataga agggaatcat cttgttcaaa 180
gcaagaatag gagtctcata gatgacaaaa ttcgggaaat gcaagcattt tttggattga 240
cagtgactgg aaaactggac tcaaacaccc ttgagatcat gaagacaccc aggtgtgggg 300
tgcctgatgt gggccagtat ggctacaccc tccctgggtg gagaaaatac aacctcacct 360
acagaataat aaactatact ccggatatgg cacgagctgc tgtggatgag gctatccaag 420
aaggtttaga agtgtggagc aaagtcactc cactaaaatt caccaagatt tcaaagggga 480
ttgcagacat catgattgcc tttaggactc gagtccatgg tcggtgtcct cgctattttg 540
atggtccctt gggagtgctt ggccatgcct ttcctcctgg tccgggtctg ggtggtgaca 600
ctcattttga tgaggatgaa aactggacca aggatggagc aggattcaac ttgtttcttg 660
tggctgctca tgaatttggt catgcactgg ggctctctca ctccaatgat caaacagcct 720
tgatgttccc aaattatgtc tccctggatc ccagaaaata cccactttct caggatgata 780
tcaatggaat ccagtccatc tatggaggtc tgcctaaggt acctgctaag ccaaaggaac 840
ccactatacc ccatgcctgt gaccctgact tgacttttga cgctatcaca actttccgca 900
gagaagtaat gttctttaaa ggcaggcacc tatggaggat ctattatgat atcacggatg 960
ttgagtttga attaattgct tcattctggc catctctgcc agctgatctg caagctgcat 1020
acgagaaccc cagagataag attctggttt ttaaagatga aaacttctgg atgatcagag 1080
gatatgctgt cttgccagat tatcccaaat ccatccatac attaggtttt ccaggacgtg 1140
tgaagaaaat agatgcagcc gtctgtgata agaccacaag aaaaacctac ttctttgtgg 1200
gcatttggtg ctggaggttt gatgaaatga cccaaaccat ggacaaagga ttcccgcaga 1260
gagtggtaaa acactttcct ggaatcagta tccgtgttga tgctgctttc cagtacaaag 1320
gattcttctt tttcagccgt ggatcaaagc aatttgaata caacattaag acaaagaata 1380
ttacccgaat catgagaact aatacttggt ttcaatgcaa agaaccaaag aactcctcat 1440
ttggttttga tatcaacaag gaaaaagcac attcaggagg cataaagata ttgtatcata 1500
agagtttaag cttgtttatt tttggtattg ttcatttgct gaaaaacact tctatttatc 1560
aataaattca tagacctaaa ataaacctca acaggtcttt taatataaat tctgcttcaa 1620
aatagaa 1627

<210> 29
<211> 1403
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5388674CB1

<400> 29
cgcagtgcga agggacgcgg tgcgcatgcg cgtgagggct gccgcggcca ggcccagaca 60
tgtccgtcct tgtaagttaa aagcttccat gggagccttc cttcctaatc aagatgcaaa 120
taatacggca ctccgaacag acactaaaaa cagctctcat ctcaaagaac ccagtgcttg 180
tatcacagta tgagaaatta gatgctgggg aacaacgttt aatgaatgaa gccttccagc 240
cagccagtga tctctttgga cccattacct tgcattctcc atcagattgg atcacctccc 300
accctgaggc tccccaagac tttgaacagt tcttcagtga tccttacaga aagacaccct 360
ctccaaacaa acgcagcatt tatatacagt ccattggtaa atactggctc tctaggaaac 420
accagaatta tcagtgaaga atatattaaa tggctcacgg gctactgtaa agcatatttc 480
tatggcttga gagtaaaact cctagaacca gttcctgttt ctgtaacaag atgttccttt 540
agagtcaatg agaacacaca caacctacaa attcatgcag gggacatcct gaagttcttg 600
aaaaagaaga aacctgaaga tgccttctgt gttgtgggaa taacaatgat tgatctttac 660
ccaagagact cgtggaattt tgtctttgga caggcctctt tgacagatgg tgtggggata 720

ttcagctttg ccaggtatgg cagtgatttt tatagcatgc actataaagg caaagtgaag 780
aagctcaaga aaacatcttc aagtgactat tcaattttcg acaactatta tattccagaa 840
ataactagtg ttttactact tcgatcctgt aagactttaa cccatgagat cggacacata 900
tttggactgc gacactgcca gtggcttgca tgcctcatgc aaggctccaa ccacttggaa 960
gaagctgacc ggcgccctct aaacctttgc cctatctgtt tgcacaagtt gcagtgtgct 1020
gttggcttca gcattgtaga aagatacaaa gcactggtga ggtggattga tgatgaatct 1080
tctgacacac ctggagcaac tccagaacac agtcacgagg ataatgggaa tttaccgaaa 1140
cccgtggaag cctttaagga atggaaagag tggataataa aatgcctggc tgttctccaa 1200
aaatgaggac cttcaaatag gagtgattga aataaataac tacttgcatg ttatgctttc 1260
atttgggtgg aatacttcat tggaataaac tactgatctt gtgctgtgtc aaagtaacag 1320
actagaacct tctttcaagt acctgaattg aaatgaaaca ttttgaataa taaaaactct 1380
agaaactcaa aaaaaaaaaa aaa 1403

<210> 30
<211> 2927
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1873102CB1

<400> 30
ctctggcttc gactccgtcg ctctcaattc gtcaccagga ggaagacgga gctggctgcc 60
cagcccaaag gcccatgagg ggatgcagtt atgggctctg tcgccgtgga ttgttatttt 120
gtgtcagtaa gtaatccata aagtgccaac atgggaaaga aacggacaaa gggaaaaact 180
gttccaatcg atgattcctc tgaaacttta gaacctgtgt gcagacacat tagaaaagga 240
ttggaacaag gtaatttgaa aaaggcttta gtgaatgtgg aatggaatat ctgccaagac 300
tgtaagactg acaataaagt gaaagataaa gctgaagaag aaacagaaga aaagccttca 360
gtttggctgt gtcttaaatg tggccatcag ggctgtggca gaaattctca ggagcagcat 420
gccttgaagc actatctgac gccaagatct gaacctcact gtctggttct tagtttggac 480
aactggagtg tatggtttta cgtatgtgat aatgaggtcc agtattgtag ttcaaaccag 540
ttgggtcaag tggttgatta tgtcagaaaa caagccagca ttacaactcc aaagccagca 600
gagaaagata atggaaatat tgaacttgaa aataaaaaat tagaaaaaga gagtaagaat 660
gaacaagaga gagaaaagaa ggaaaacatg gctaaagaga atcctcccat gaattctcct 720
tgccaaataa ccgtgaaagg actcagtaat ttgggaaaca catgtttctt caatgcagtt 780
atgcagaact tgtcacaaac accagtgctt agagaactac taaaagaagt gaaaatgtct 840
ggaacaattg taaaaattga accacctgat ttggcattaa cagaaccatt agaaataaac 900
cttgagcctc caggccctct tactttagcc atgagccagt ttcttaatga gatgcaagag 960
accaaaaagg gggttgtgac accgaaagaa ctcttttctc aggtctgtaa aaaagcagtg 1020
cggtttaaag gctatcagca gcaagacagc caggagctgc ttcgctactt attggatggg 1080
atgagagcag aagaacacca aagagtgagt aaaggaatac ttaaagcatt tggtaattct 1140
actgaaaagt tggatgaaga actaaaaaat aaagttaaag attatgagaa gaaaaaatca 1200
atgccaagtt ttgttgaccg catctttggt ggtgaactaa ctagtatgat catgtgtgat 1260
caatgcagaa ctgtctcctt ggttcatgaa tctttccttg atttgtccct cccagtttta 1320
gatgatcaga gtggtaagaa aagtgtaaat gataaaaatc tgaaaaagac agtggaggat 1380
gaagatcaag atagtgagga agaaaaagat aacgacagtt acataaaaga gagaagtgat 1440
attccttctg gaacaagtaa gcacttacag aaaaaagcaa agaaacaagc caaaaagcaa 1500
gccaagaacc aacgaagaca acaaaaaatt caaggaaaag ttcttcattt aaatgatatt 1560
tgtactattg accatcctga agacagtgaa tatgaagctg aaatgtcact tcaaggagaa 1620
gtaaatatta aatccaacca tatttcacaa gagggtgtta tgcataaaga atattgtgtc 1680
aaccagaaag atttgaatgg ccaagcaaaa atgatcgaaa gtgtaactga caatcaaaaa 1740
tccacagagg aagtagatat gaaaaatatc aacatggata atgatctgga ggttttaaca 1800
tcttctccca ctaggaattt aaatggtgcc tacctaacgg aagggagcaa tggagaagtg 1860

gacatttcca atggtttcaa aaacctaaat ttgaatgctg ctcttcatcc tgatgaaata 1920
aatatagaga ttctgaatga tagtcatact cctggaacaa aggtgtatga ggttgtaaat 1980
gaagatccag aaactgcttt ctgtactctt gcaaacaggg aagttttcaa tactgatgag 2040
tgttcaatcc aacattgttt atatcagttc acccgtaatg agaaacttcg agatgcgaat 2100
aaactgcttt gtgaagtatg cacacggaga cagtgtaatg gaccaaaggc aaatataaaa 2160
ggtgaaagga agcatgttta caccaatgcc aaaaagcaga tgctaatttc tcttgctcct 2220
cctgttctta ctcttcattt aaagagattt cagcaggctg gttttaacct acgcaaagtt 2280
aacaaacaca taaagtttcc ggaaatctta gatttggctc ctttttgcac ccttaaatgt 2340
aagaatgttg cagaagaaaa tacaagggta ctctattcct tatatggagt tgttgaacac 2400
agtggtacta tgaggtcggg gcattacact gcctatgcca aggcaagaac cgcaaatagt 2460
catctctcta atcttgttct tcacggtgat attccacaag attttgaaat ggaatcaaaa 2520
gggcagtggt ttcacatcag cgacacacat gtgcaagctg tgcctacaac taaagtacta 2580
aactcacaag cgtacctcct attttatgag agaatactgt aataatatca aaagcacttt 2640
ttctggaaac acatttatgg cttttataat ggctgaaata acgataaaaa aagactaatt 2700
aaaatcatgt tcacttaaca ttaaatacat gccagaagaa atcatgttta tttaaatatt 2760
gaagggaaaa atacctaaaa atgtacaaag gttttatatt gtcatagtgg tttttattcc 2820
tgctttgttt ctggaaagga aatcctgaat tacttaagta ctttgtgttt aatatatctg 2880
ggtgatggat cacaacacat caataaactg acttacccta aaaaaaa 2927

<210> 31
<211> 1526
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1920734CB1

<400> 31
ctcgagccgg gccgcctaag tcccacagag acgggagtcg ggtgggatcc caggctgggc 60
cccgcggcgg ctggattctc ttccctggcc aagtctctga gatcttctcc cagggcgatg 120
caaagctact cgctaccagc ttggacctgt ctgcagtatc tcctctggga cctgccatgc 180
tgaggaccca ttctcacctc tgagggactc ctgtcctagg actaaggtgg agcctgggcc 240
atggtacagc tggctcctgc ggcagccatg gacgaggtca cctttaggag cgacactgtg 300
ctgtcagatg tccacctcta taccccgaac catagacatc tcatggtacg gctgaacagc 360
gtggggcagc cagttttcct gtcccaattc aagcttctat ggagccaaga ctcttggaca 420
gattcaggag ccaagggtgg cagtcacaga gatgttcaca caaaggagcc tccttctgct 480
gagacaggca gcacagggtc ccctccagga agtggccatg gtaatgaggg tttctccctc 540
caggccggga ctgacaccac tggccaggaa gtggctgaag ctcagctgga tgaggatggg 600
gatttggacg tggtgagaag accacgagcc gcctctgatt ccaacccagc agggcctctg 660
agagacaagg tacatcccat gattctagca caggaagaag acgacgtcct gggagaggaa 720
gcacaaggca gcccgcacga tatcatcaga atagagcaca ccatggccac gcccctggag 780
gatgttggca agcaggtgtg gcggggcgcc ctgctcctgg cagactacat cctgttccga 840
caggacctct tccgaggatg tacagcgctg gagctcgggg ccggcacggg gctcgctagc 900
atcatcgcag ccaccatggc acggaccgtt tattgtacag atgtcggtgc agatctcttg 960
tccatgtgcc agcgaaacat tgccctcaac agccacctgg ctgccactgg aggtggtata 1020
gttagggtca aagaactgga ctggctgaag gacgacctct gcacagatcc caaggtcccc 1080
ttcagttggt cacaagagga aatttctgac ctgtacgatc acaccaccat cctgtttgca 1140
gccgaagtgt tttacgacga cgacttgact gatgctgtgt ttaaaacgct ctcccgactc 1200
gcccacagat tgaaaaatgc ctgcacagcc atactgtcgg tggagaagag gctcaacttc 1260
acactgagac acttggacgt cacatgtgaa gcctacgatc acttccgctc ctgcctgcac 1320
gcgctggagc agctcacaga tggcaagctg cgcttcgtgg tggagcccgt ggaggcctcc 1380
ttcccacagc tcctggttta cgagcgcctc cagcagctgg agctctggaa gatcatcgca 1440
gaaccagtaa catgacccat cgcctccacc aggcgcggcg tctcgactgt tcttagagtg 1500

tatttctagt aaaatcagaa gctcac 1526

<210> 32
<211> 2390
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2396858CB1

<400> 32
ggaagttgag cggcggcaag aaataatggc ggcagctacg ggggatcctg gactctctaa 60
actgcagttt gccccttta gtagtgcctt ggatgttggg ttttggcatg agttgaccca 120
gaagaagctg aacgagtatc ggctggatga agctcccaag gacattaagg gttattacta 180
caatggtgac tctgctgggc tgccagctcg cttaacattg gagttcagtg cttttgacat 240
gagtgctccc accccagccc gttgctgccc agctattgga acactgtata acaccaacac 300
actcgagtct ttcaagactg cagataagaa gctccttttg gaacaagcag caaatgagat 360
atgggaatcc ataaaatcag gcactgctct tgaaaaccct gtactcctca acaagttcct 420
cctcttgaca tttgcagatc taaagaagta ccacttctac tattggtttt gctatcctgc 480
cctctgtctt ccagagagtt tacctctcat tcaggggcca gtgggtttgg atcaaaggtt 540
ttcactaaaa cagattgaag cactagagtg tgcatatgat aatctttgtc aaacagaagg 600
agtcacagct cttccttact tcttaatcaa gtatgatgag aacatggtgc tggtttcctt 660
gcttaaacac tacagtgatt tcttccaagg tcaaaggacg aagataacaa ttggtgtata 720
tgatccctgt aacttagccc agtaccctgg atggcctttg aggaattttt tggtcctagc 780
agcccacaga tggagtagca gtttccagtc tgttgaagtt gtttgcttcc gtgaccgtac 840
catgcagggg gcgagagacg ttgcccacag catcatcttc gaagtgaagc ttccagaaat 900
ggcatttagc ccagattgtc ctaaagcagt tggatgggaa aagaaccaga aaggaggcat 960
gggaccaagg atggtgaacc tcagtgaatg tatggaccct aaaaggttag ctgagtcatc 1020
agtggatcta aatctcaaac tgatgtgttg gagattggtt cctactttag acttggacaa 1080
ggttgtgtct gtcaaatgtc tgctgcttgg agccggcacc ttgggttgca atgtagctag 1140
gacgttgatg ggttggggcg tgagacacat cacatttgtg gacaatgcca agatctccta 1200
ctccaatcct gtgaggcagc ctctctatga gtttgaagat tgcctagggg gtggtaagcc 1260
caaggctctg gcagcagcgg accggctcca gaaaatattc cccggtgtga atgccagagg 1320
attcaacatg agcataccta tgcctgggca tccagtgaac ttctccagtg tcactctgga 1380
gcaagcccgc agagatgtgg agcaactgga gcagctcatc gaaagccatg atgtcgtctt 1440
cctattgatg gacaccaggg agagccggtg gcttcctgcc gtcattgctg caagcaagag 1500
aaagctggtc atcaatgctg ctttgggatt tgacacattt gttgtcatga gacatggtct 1560
gaagaaacca aagcagcaag gagctgggga cttgtgtcca aaccaccctg tggcatctgc 1620
tgacctcctg ggctcatcgc tttttgccaa catccctggt tacaagcttg gctgctactt 1680
ctgcaatgat gtggtggccc caggagattc aaccagagac cggaccttgg accagcagtg 1740
cactgtgagt cgtccaggac tggccgtgat tgcaggagcc ctggccgtgg aattgatggt 1800
atctgttttg cagcatccag aaggggggcta tgccattgcc agcagcagtg acgatcggat 1860
gaatgagcct ccaacctctc ttgggcttgt gcctcaccag atccggggat ttctttcacg 1920
gtttgataat gtccttcccg tcagcctggc atttgacaaa tgtacagctt gttcttccaa 1980
agttcttgat caatatgaac gagaaggatt taacttccta gccaaggtgt ttaattcttc 2040
acattccttc ttagaagact tgactggtct tacattgctg catcaagaaa cccaagctgc 2100
tgagatctgg gacatgagcg atgatgagac catctgagat ggccccgctg tggggctgac 2160
ttctccctgg ccgcctgctg aggagctctc catcgccaga gcaggactgc tgaccccagg 2220
cctggtgatt ctgggcccct cctccatacc ccgaggtctg ggattccccc ctctgctgcc 2280
caggagtggc cagtgttcgg cgttgctcgg gattcaagat accaccagtt cagagctaaa 2340
taataacctt ggccttggcc ttgctattga cctggaaaaa aaaaaaaaaa 2390

<210> 33
<211> 1241
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2634725CB1

<400> 33
aacaggactt acagctccaa tgccctcaca gaaccccgcg ggcatttgcg gcgccttccc 60
tccgcctctc tgcttcttgt tccctcaccc ctgcttgcca cggcgtctgt tccctctcaa 120
ctcgcctggc agatctcagc atgcccctgc ttcctctagg tggctgctat tgccagagcc 180
cactgagtag acgcggatta cccgtttgtc agccccgggc atctgcgctg ttccagagtt 240
ttctagtcca atgacccgtg gggagcgtct cgaatgacgc tgccctcgaa gcagcccggc 300
tctcagcccc gtcctgccct ctccccaggc actggcgcgc tgatcctgca aaagggtgag 360
atccgcgtca tcaaccagac cacctgcgag aacctcctgc cgcagcagat cacgccgcgc 420
atgatgtgcg tgggcttcct cagcggcggc gtggactcct gccaggtggc ccccggggca 480
ggagggcggc aggtgggccc cgggagaggc gggactgggg actcaccggc agggcttgtc 540
tccgcccagg gtgattccgg gggacccctg tccagcgtgg aggcggatgg gcggatcttc 600
caggccggtg tggtgagctg gggagacggc tgcgctcaga ggaacaagcc aggcgtgtac 660
acaaggctcc ctctgtttcg ggactggatc aaagagaaca ctggggtata ggggccgggg 720
ccacccaaat gtgtacacct gcggggccac ccatcgtcca ccccagtgtg cacgcctgca 780
ggctggagac tggaccgctg actgcaccag cgcccccaga acatacactg tgaactcaat 840
ctccagggct ccaaatctgc ctagaaaacc tctcgcttcc tcagcctcca aagtggagct 900
gggaggtaga aggggaggac actggtggtt ctactgaccc aactgggggc aaaggtttga 960
agacacagcc tcccccgcca gccccaagct gggccgaggc gcgtttgtgt atatctgcct 1020
cccctgtctg taaggagcag cgggaacgga gcttcggagc ctcctcagtg aaggtggtgg 1080
ggctgccgga tctgggctgt ggggcccttg ggccacgctc ttgaggaagc ccaggctcgg 1140
aggaccctgg aaaacagacg ggtctgagac tgaaattgtt ttaccagctc ccagggtgga 1200
cttcagtgtg tgtatttgtg taaatgagta aaacatttta t 1241

<210> 34
<211> 2079
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2643110CB1

<400> 34
ggtgcctgag ccggcgggtc ccctgtgtcc gccgcggctg tcgtcccccg ctcccgccac 60
ttccggggtc gcagtcccgg gcatggagcc gcgaccgtga ggcgccgctg gacccgggac 120
gacctgccca gtccggccgc cgccccacgt cccggtctgt gtcccacgcc tgcagctgga 180
atggaggctc tctggaccct ttagaaggca cccctgccct cctgaggtca gctgagcggt 240
taatgcggaa ggttaagaaa ctgcgcctgg acaaggagaa caccggaagt tggagaagct 300
tctcgctgaa ttccgagggg gctgagagga tggccaccac cgggacccca acggccgacc 360
gatgcgacgc agccgccaca gatgacccgg ccgcccgctt ccaggtgcag aagcactcgt 420
gggacgggct ccggagcatc atccacggca gccgcaagta ctcgggcctc attgtcaaca 480
aggcgcccca cgacttccag tttgtgcaga agacggatga gtctgggccc cactcccacc 540
gcctctacta cctgggaatg ccatatggca gccgagagaa ctccctcctc tactctgaga 600
ttcccaagaa ggtccggaaa gaggctctgc tgctcctgtc ctggaagcag atgctggatc 660
atttccaggc cacgccccac catggggtct actctcggga ggaggagctg ctgagggagc 720

ggaaacgcct gggggtcttc ggcatcacct cctacgactt ccacagcgag agtggcctct 780
tcctcttcca ggccagcaac agcctcttcc actgccgcga cggcggcaag aacggcttca 840
tggtgtcccc tatgaaaccg ctggaaatca agacccagtg ctcagggccc cggatggacc 900
ccaaaatctg ccctgccgac cctgacttct tctccttcat caataacagc gacctgtggg 960
tggccaacat cgagacaggc gaggagcggc ggctgacctt ctgccaccaa ggtttatcca 1020
atgtcctgga tgaccccaag tctgcgggtg tggccacctt cgtcatacag gaagagttcg 1080
accgcttcac tgggtactgg tggtgcccca cagcctcctg ggaaggttca gagggcctca 1140
agacgctgcg aatcctgtat gaggaagtcg atgagtccga ggtggaggtc attcacgtcc 1200
cctctcctgc gctagaagaa aggaagacgg actcgtatcg gtaccccagg acaggcagca 1260
agaatcccaa gattgccttg aaactggctg agttccagac tgacagccag ggcaagatcg 1320
tctcgaccca ggagaaggag ctggtgcagc ccttcagctc gctgttcccg aaggtggagt 1380
acatcgccag ggccgggtgg acccgggatg gcaaatacgc ctgggccatg ttcctggacc 1440
ggccccagca gtggctccag ctcgtcctcc tccccccggc cctgttcatc ccgagcacag 1500
agaatgagga gcagcggcta gcctctgcca gagctgtccc caggaatgtc cagccgtatg 1560
tggtgtacga ggaggtcacc aacgtctgga tcaatgttca tgacatcttc tatcccttcc 1620
cccaatcaga gggagaggac gagctctgct ttctccgcgc caatgaatgc aagaccggct 1680
tctgccattt gtacaaagtc accgccgttt taaaatccca gggctacgat tggagtgagc 1740
ccttcagccc cggggaaggt gagcagagcc tgacgaatgc tgtcgactca tcgcgttagt 1800
cacgtgtggt tcaatatgct gtttgttcat tggtcggccc ccccactcag ccagcacacc 1860
ctgcgggaga aggaacaggg atcggcagga agccagcctt ccccagtgac tgcatgatct 1920
ggcagggctt agagcaccca actgttggct tattcaggca gcagatttac tgagcacctc 1980
ccctgtgcca ggcccttagc acaaccaggg gttggccacc tacggcccac aggtcaaatc 2040
cggcccacca cctgtgttca taaataaagt tttattggc 2079

<210> 35
<211> 1731
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2701396CB1

<400> 35
cttaatgact agaattcagg ttccaaggag aagcccacaa ggctaagggt attggatata 60
acggaaagtg gaagctatac ctgacttcca gagaatgtgg accggatata agatcttaat 120
cttctcttat cttactacag aaatctggat ggagaagcag tatttatctc aaagagaagt 180
ggacctagag gcttatttca ctaggaatca caccgttttg caaggtactc gattcaaaag 240
agccattttc caagggcaat actgtagaaa ttttggctgt tgtgaagaca gagatgatgg 300
ctgtgtcact gagttctatg cggcgaatgc gttgtgctac tgtgataaat tctgtgacag 360
agaaaattct gattgctgtc ctgactacaa gtcctttgc cgtgaagaga aagaatggcc 420
tcctcacaca cagccttggt atccagaagg ttgcttcaaa gatggtcaac attatgaaga 480
gggatcagta attaaagaaa actgcaactc ctgcacatgc tcaggacagc aatggaaatg 540
ttcccagcat gtatgccttg ttcgttcaga attaattgaa caggtcaata aaggagacta 600
tggatggaca gcacagaatt acagccaatt ttggggaatg actttagaag atggttttaa 660
atttcgcctt ggcactttgc cacctagtcc catgctcctg agcatgaatg aaatgacagc 720
ttctttacct gcaacaactg atcttccaga gtttttgttg cttcttataa atggcctgga 780
tggactcatg gcccattgga tcaaaaaaat ttgtgctgca tcctgggcat tttccactgc 840
aagtgtggct gctgaccgaa tagcaattca gtctaagggt cgatacacgg ccaatctatc 900
ccctcagaat ttgatctctt gctgtgccaa gaaccgtcat ggatgcaata gtggaagcat 960
cgatagggct tggtggtacc tgagaaaacg tggactggta tcccacgcat gctacccact 1020
tttcaaagac caaaatgcta ccaacaatgg atgtgccatg gcaagcaggt ctgatgggcg 1080
aggaaaacgg catgccacga agccatgtcc caacaacgta gaaaaatcta acaggatcta 1140
tcaatgttct cctccataca gagtctcttc caacgaaact gagataatga aagaaatcat 1200

gcaaaatgga ccagttcaag ccataatgca agtccgtgaa gatttcttcc attataagac 1260
agggatatac agacatgtta ccagcacaaa taaagaatca gaaaaatatc gaaagcttca 1320
gacacatgca gtcaaactca ctggatgggg cacactgaga ggagcacaag ggcagaaaga 1380
aaaattttgg attgctgcca attcctgggg aaagtcatgg ggagagaatg gctatttcag 1440
gattcttcga ggagtaaatg agtccgacat tgaaaagttg attatcgcag cttggggcca 1500
actgacgagt tctgatgaac cataacatat cattaaattt ccataaggtc atgcctttaa 1560
gtaacccccc aaattgaagt ttagcaatat gacattcttg gtgacagtgg aatctttgtc 1620
tcttcaccgt gttaacataa tctatctatt ttcttatttt cccctctggt ctatgcttct 1680
gcttccttca tattactgag cattaacaac accaataaag gacagcagag t 1731

<210> 36
<211> 1081
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3134404CB1

<400> 36
ggaaaaaggt gtctagctcc tttctgctta aaaaagcaca gggagatcgc gggcagcttt 60
gcagtcgctg ccttctcgcg cctgaccatg cacccctgca tcttcctgct gggccacagg 120
cgagcgcttt atttctggag ctgagggcta aaacttttct gacttttctt ctcctcaaca 180
tctgaatcat gccatgtgcc cagaggagct ggcttgcaaa cctttccgtg gtggctcagc 240
tccttaactt tggggcgctt tgctatggga gacagcttca gccaggcccg gttcgcttcc 300
cggacaggag gcaagagcat tttatcaagg gcctgccaga ataccacgtg gtgggtccag 360
tccgagtaga tgccagtggg catttttgt catatggctt gcactatccc atcacgagca 420
gcaggaggaa gagagatttg gatggctcag aggactgggt gtactacaga atttctcacg 480
aggagaagga cctgtttttt aacttgacgg tcaatcaagg atttctttcc aatagctaca 540
tcatggagaa gagatatggg aacctctccc atgttaagat gatggcttcc tctgcccccc 600
tctgccatct cagtggcacg gttctacagc agggcaccag agttgggacg gcagccctca 660
gtgcctgcca tggactgact ggatttttcc aactaccaca tggagacttt ttcattgaac 720
ccgtgaagaa gcatccactg gttgagggag ggtaccaccc gcacatcgtt tacaggaggc 780
agaaagttcc agaaaccaag gagccaacct gtggattaaa gggtattgtg actcacatgt 840
cctcctgggt tgaagaatct gttttgttct tttggtagtt ttattaaaac atgacctatt 900
cttactcaag tctcttatct cctctgtatt ctttttttct taatatcttc atgacattca 960
aatctcttct gtattctctt gccagaaagt gtacattctt tttgcttgta taaacccttt 1020
cacttgtcaa tactcgagtc tctgcatatt cttaatgaga aggaaataaa agcatctttg 1080
c 1081

<210> 37
<211> 1019
<212> PRT
<213> Homo sapiens

<300>
<308> GenBank ID No: g746413

<400> 37
Met Gly Ser Lys Arg Gly Ile Ser Ser Arg His His Ser Leu Ser
1 5 10 15
Ser Tyr Glu Ile Met Phe Ala Ala Leu Phe Ala Ile Leu Val Val
20 25 30

```
Leu Cys Ala Gly Leu Ile Ala Val Ser Cys Leu Thr Ile Lys Glu
                35                  40                  45
Ser Gln Arg Gly Ala Ala Leu Gly Gln Ser His Glu Ala Arg Ala
            50                  55                  60
Thr Phe Lys Ile Thr Ser Gly Val Thr Tyr Asn Pro Asn Leu Gln
        65                  70                  75
Asp Lys Leu Ser Val Asp Phe Lys Val Leu Ala Phe Asp Leu Gln
                80                  85                  90
Gln Met Ile Asp Glu Ile Phe Leu Ser Ser Asn Leu Lys Asn Glu
                95                  100                 105
Tyr Lys Asn Ser Arg Val Leu Gln Phe Glu Asn Gly Ser Ile Ile
            110                 115                 120
Val Val Phe Asp Leu Phe Phe Ala Gln Trp Val Ser Asp Gln Asn
        125                 130                 135
Val Lys Glu Glu Leu Ile Gln Gly Leu Glu Ala Asn Lys Ser Ser
    140                 145                 150
Gln Leu Val Thr Phe His Ile Asp Leu Asn Ser Val Asp Ile Leu
155                 160                 165
Asp Lys Leu Thr Thr Thr Ser His Leu Ala Thr Pro Gly Asn Val
            170                 175                 180
Ser Ile Glu Cys Leu Pro Gly Ser Ser Pro Cys Thr Asp Ala Leu
        185                 190                 195
Thr Cys Ile Lys Ala Asp Leu Phe Cys Asp Gly Glu Val Asn Cys
    200                 205                 210
Pro Asp Gly Ser Asp Glu Asp Asn Lys Met Cys Ala Thr Val Cys
215                 220                 225
Asp Gly Arg Phe Leu Leu Thr Gly Ser Ser Gly Ser Phe Gln Ala
            230                 235                 240
Thr His Tyr Pro Lys Pro Ser Glu Thr Ser Val Val Cys Gln Trp
        245                 250                 255
Ile Ile Arg Val Asn Gln Gly Leu Ser Ile Lys Leu Ser Phe Asp
    260                 265                 270
Asp Phe Asn Thr Tyr Tyr Thr Asp Ile Leu Asp Ile Tyr Glu Gly
275                 280                 285
Val Gly Ser Ser Lys Ile Leu Arg Ala Ser Ile Trp Glu Thr Asn
            290                 295                 300
Pro Gly Thr Ile Arg Ile Phe Ser Asn Gln Val Thr Ala Thr Phe
        305                 310                 315
Leu Ile Glu Ser Asp Glu Ser Asp Tyr Val Gly Phe Asn Ala Thr
    320                 325                 330
Tyr Thr Ala Phe Asn Ser Ser Glu Leu Asn Asn Tyr Glu Lys Ile
                335                 340                 345
Asn Cys Asn Phe Glu Asp Gly Phe Cys Phe Trp Val Gln Asp Leu
            350                 355                 360
Asn Asp Asp Asn Glu Trp Glu Arg Ile Gln Gly Ser Thr Phe Ser
        365                 370                 375
Pro Phe Thr Gly Pro Asn Phe Asp His Thr Phe Gly Asn Ala Ser
    380                 385                 390
Gly Phe Tyr Ile Ser Thr Pro Thr Gly Pro Gly Gly Arg Gln Glu
395                 400                 405
Arg Val Gly Leu Leu Ser Leu Pro Leu Asp Pro Thr Leu Glu Pro
            410                 415                 420
Ala Cys Leu Ser Phe Trp Tyr His Met Tyr Gly Glu Asn Val His
        425                 430                 435
Lys Leu Ser Ile Asn Ile Ser Asn Asp Gln Asn Met Glu Lys Thr
```

```
                    440                 445                 450
Val Phe Gln Lys Glu Gly Asn Tyr Gly Asp Asn Trp Asn Tyr Gly
                    455                 460                 465
Gln Val Thr Leu Asn Glu Thr Val Lys Phe Lys Val Ala Phe Asn
                    470                 475                 480
Ala Phe Lys Asn Lys Ile Leu Ser Asp Ile Ala Leu Asp Asp Ile
                    485                 490                 495
Ser Leu Thr Tyr Gly Ile Cys Asn Gly Ser Leu Tyr Pro Glu Pro
                    500                 505                 510
Thr Leu Val Pro Thr Pro Pro Pro Glu Leu Pro Thr Asp Cys Gly
                    515                 520                 525
Gly Pro Phe Glu Leu Trp Glu Pro Asn Thr Thr Phe Ser Ser Thr
                    530                 535                 540
Asn Phe Pro Asn Ser Tyr Pro Asn Leu Ala Phe Cys Val Trp Ile
                    545                 550                 555
Leu Asn Ala Gln Lys Gly Lys Asn Ile Gln Leu His Phe Gln Glu
                    560                 565                 570
Phe Asp Leu Glu Asn Ile Asn Asp Val Val Glu Ile Arg Asp Gly
                    575                 580                 585
Glu Glu Ala Asp Ser Leu Leu Leu Ala Val Tyr Thr Gly Pro Gly
                    590                 595                 600
Pro Val Lys Asp Val Phe Ser Thr Thr Asn Arg Met Thr Val Leu
                    605                 610                 615
Leu Ile Thr Asn Asp Val Leu Ala Arg Gly Gly Phe Lys Ala Asn
                    620                 625                 630
Phe Thr Thr Gly Tyr His Leu Gly Ile Pro Glu Pro Cys Lys Ala
                    635                 640                 645
Asp His Phe Gln Cys Lys Asn Gly Glu Cys Val Pro Leu Val Asn
                    650                 655                 660
Leu Cys Asp Gly His Leu His Cys Glu Asp Gly Ser Asp Glu Ala
                    665                 670                 675
Asp Cys Val Arg Phe Phe Asn Gly Thr Thr Asn Asn Asn Gly Leu
                    680                 685                 690
Val Arg Phe Arg Ile Gln Ser Ile Trp His Thr Ala Cys Ala Glu
                    695                 700                 705
Asn Trp Thr Thr Gln Ile Ser Asn Asp Val Cys Gln Leu Leu Gly
                    710                 715                 720
Leu Gly Ser Gly Asn Ser Ser Lys Pro Ile Phe Ser Thr Asp Gly
                    725                 730                 735
Gly Pro Phe Val Lys Leu Asn Thr Ala Pro Asp Gly His Leu Ile
                    740                 745                 750
Leu Thr Pro Ser Gln Gln Cys Leu Gln Asp Ser Leu Ile Arg Leu
                    755                 760                 765
Gln Cys Asn His Lys Ser Cys Gly Lys Lys Leu Ala Ala Gln Asp
                    770                 775                 780
Ile Thr Pro Lys Ile Val Gly Gly Ser Asn Ala Lys Glu Gly Ala
                    785                 790                 795
Trp Pro Trp Val Val Gly Leu Tyr Tyr Gly Gly Arg Leu Leu Cys
                    800                 805                 810
Gly Ala Ser Leu Val Ser Ser Asp Trp Leu Val Ser Ala Ala His
                    815                 820                 825
Cys Val Tyr Gly Arg Asn Leu Glu Pro Ser Lys Trp Thr Ala Ile
                    830                 835                 840
Leu Gly Leu His Met Lys Ser Asn Leu Thr Ser Pro Gln Thr Val
                    845                 850                 855
```

```
Pro Arg Leu Ile Asp Glu Ile Val Ile Asn Pro His Tyr Asn Arg
                860                 865                 870
Arg Arg Lys Asp Asn Asp Ile Ala Met Met His Leu Glu Phe Lys
                875                 880                 885
Val Asn Tyr Thr Asp Tyr Ile Gln Pro Ile Cys Leu Pro Glu Glu
                890                 895                 900
Asn Gln Val Phe Pro Pro Gly Arg Asn Cys Ser Ile Ala Gly Trp
                905                 910                 915
Gly Thr Val Val Tyr Gln Gly Thr Thr Ala Asn Ile Leu Gln Glu
                920                 925                 930
Ala Asp Val Pro Leu Leu Ser Asn Glu Arg Cys Gln Gln Gln Met
                935                 940                 945
Pro Glu Tyr Asn Ile Thr Glu Asn Met Ile Cys Ala Gly Tyr Glu
                950                 955                 960
Glu Gly Gly Ile Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu
                965                 970                 975
Met Cys Gln Glu Asn Asn Arg Trp Phe Leu Ala Gly Val Thr Ser
                980                 985                 990
Phe Gly Tyr Lys Cys Ala Leu Pro Asn Arg Pro Gly Val Tyr Ala
                995                1000                1005
Arg Val Ser Arg Phe Thr Glu Trp Ile Gln Ser Phe Leu His
               1010                1015
```

<210> 38
<211> 535
<212> PRT
<213> Methanococcus jannaschii

<300>
<308> GenBank ID No: g2826367

<400> 38

```
Met Ile Cys Leu Gly Leu Glu Gly Thr Ala Glu Lys Thr Gly Val
  1               5                  10                  15
Gly Ile Val Thr Ser Asp Gly Glu Val Leu Phe Asn Lys Thr Ile
                20                  25                  30
Met Tyr Lys Pro Pro Lys Gln Gly Ile Asn Pro Arg Glu Ala Ala
                35                  40                  45
Asp His His Ala Glu Thr Phe Pro Lys Leu Ile Lys Glu Ala Phe
                50                  55                  60
Glu Val Val Asp Lys Asn Glu Ile Asp Leu Ile Ala Phe Ser Gln
                65                  70                  75
Gly Pro Gly Leu Gly Pro Ser Leu Arg Val Thr Ala Thr Val Ala
                80                  85                  90
Arg Thr Leu Ser Leu Thr Leu Lys Lys Pro Ile Ile Gly Val Asn
                95                 100                 105
His Cys Ile Ala His Ile Glu Ile Gly Lys Leu Thr Thr Glu Ala
               110                 115                 120
Glu Asp Pro Leu Thr Leu Tyr Val Ser Gly Gly Asn Thr Gln Val
               125                 130                 135
Ile Ala Tyr Val Ser Lys Lys Tyr Arg Val Phe Gly Glu Thr Leu
               140                 145                 150
Asp Ile Ala Val Gly Asn Cys Leu Asp Gln Phe Ala Arg Tyr Val
               155                 160                 165
```

```
Asn Leu Pro His Pro Gly Gly Pro Tyr Ile Glu Glu Leu Ala Arg
                170                 175                 180
Lys Gly Lys Lys Leu Val Asp Leu Pro Tyr Thr Val Lys Gly Met
                185                 190                 195
Asp Ile Ala Phe Ser Gly Leu Leu Thr Ala Ala Met Arg Ala Tyr
                200                 205                 210
Asp Ala Gly Glu Arg Leu Glu Asp Ile Cys Tyr Ser Leu Gln Glu
                215                 220                 225
Tyr Ala Phe Ser Met Leu Thr Glu Ile Thr Glu Arg Ala Leu Ala
                230                 235                 240
His Thr Asn Lys Gly Glu Val Met Leu Val Gly Gly Val Ala Ala
                245                 250                 255
Asn Asn Arg Leu Arg Glu Met Leu Lys Ala Met Cys Glu Gly Gln
                260                 265                 270
Asn Val Asp Phe Tyr Val Pro Pro Lys Glu Phe Cys Gly Asp Asn
                275                 280                 285
Gly Ala Met Ile Ala Trp Leu Gly Leu Leu Met His Lys Asn Gly
                290                 295                 300
Arg Trp Met Ser Leu Asp Glu Thr Lys Ile Ile Pro Asn Tyr Arg
                305                 310                 315
Thr Asp Met Val Glu Val Asn Trp Ile Lys Glu Ile Lys Gly Lys
                320                 325                 330
Lys Arg Lys Ile Pro Glu His Leu Ile Gly Lys Gly Ala Glu Ala
                335                 340                 345
Asp Ile Lys Arg Asp Ser Tyr Leu Asp Phe Asp Val Ile Ile Lys
                350                 355                 360
Glu Arg Val Lys Lys Gly Tyr Arg Asp Glu Arg Leu Asp Glu Asn
                365                 370                 375
Ile Arg Lys Ser Arg Thr Ala Arg Glu Ala Arg Tyr Leu Ala Leu
                380                 385                 390
Val Lys Asp Phe Gly Ile Pro Ala Pro Tyr Ile Phe Asp Val Asp
                395                 400                 405
Leu Asp Asn Lys Arg Ile Met Met Ser Tyr Ile Asn Gly Lys Leu
                410                 415                 420
Ala Lys Asp Val Ile Glu Asp Asn Leu Asp Ile Ala Tyr Lys Ile
                425                 430                 435
Gly Glu Ile Val Gly Lys Leu His Lys Asn Asp Val Ile His Asn
                440                 445                 450
Asp Leu Thr Thr Ser Asn Phe Ile Phe Asp Lys Asp Leu Tyr Ile
                455                 460                 465
Ile Asp Phe Gly Leu Gly Lys Ile Ser Asn Leu Asp Glu Asp Lys
                470                 475                 480
Ala Val Asp Leu Ile Val Phe Lys Lys Ala Val Leu Ser Thr His
                485                 490                 495
His Glu Lys Phe Asp Glu Ile Trp Glu Arg Phe Leu Glu Gly Tyr
                500                 505                 510
Lys Ser Val Tyr Asp Arg Trp Glu Ile Ile Leu Glu Leu Met Lys
                515                 520                 525
Asp Val Glu Arg Arg Ala Arg Tyr Val Glu
                530                 535
```

<210> 39
<211> 496
<212> PRT

<213> Homo sapiens

<300>
<308> GenBank ID No: g431321

<400> 39

```
Met Gly Arg Arg Ala Leu Leu Leu Leu Leu Ser Phe Leu Ala
  1               5                  10                  15
Pro Trp Ala Thr Ile Ala Leu Arg Pro Ala Leu Arg Ala Leu Gly
                20                  25                  30
Ser Leu His Leu Pro Thr Asn Pro Thr Ser Leu Pro Ala Val Ala
                35                  40                  45
Lys Asn Tyr Ser Val Leu Tyr Phe Gln Gln Lys Val Asp His Phe
                50                  55                  60
Gly Phe Asn Thr Val Lys Thr Phe Asn Gln Arg Tyr Leu Val Ala
                65                  70                  75
Asp Lys Tyr Trp Lys Lys Asn Gly Gly Ser Ile Leu Phe Tyr Thr
                80                  85                  90
Gly Asn Glu Gly Asp Ile Ile Trp Phe Cys Asn Asn Thr Gly Phe
                95                 100                 105
Met Trp Asp Val Ala Glu Glu Leu Lys Ala Met Leu Val Phe Ala
               110                 115                 120
Glu His Arg Tyr Tyr Gly Glu Ser Leu Pro Phe Gly Asp Asn Ser
               125                 130                 135
Phe Lys Asp Ser Arg His Leu Asn Phe Leu Thr Ser Glu Gln Ala
               140                 145                 150
Leu Ala Asp Phe Ala Glu Leu Ile Lys His Leu Lys Arg Thr Ile
               155                 160                 165
Pro Gly Ala Glu Asn Gln Pro Val Ile Ala Ile Gly Gly Ser Tyr
               170                 175                 180
Gly Gly Met Leu Ala Ala Trp Phe Arg Met Lys Tyr Pro His Met
               185                 190                 195
Val Val Gly Ala Leu Ala Ala Ser Ala Pro Ile Trp Gln Phe Glu
               200                 205                 210
Asp Leu Val Pro Cys Gly Val Phe Met Lys Ile Val Thr Thr Asp
               215                 220                 225
Phe Arg Lys Ser Gly Pro His Cys Ser Glu Ser Ile His Arg Ser
               230                 235                 240
Trp Asp Ala Ile Asn Arg Leu Ser Asn Thr Gly Ser Gly Leu Gln
               245                 250                 255
Trp Leu Thr Gly Ala Leu His Leu Cys Ser Pro Leu Thr Ser Gln
               260                 265                 270
Asp Ile Gln His Leu Lys Asp Trp Ile Ser Glu Thr Trp Val Asn
               275                 280                 285
Leu Ala Met Val Asp Tyr Pro Tyr Ala Ser Asn Phe Leu Gln Pro
               290                 295                 300
Leu Pro Ala Trp Pro Ile Lys Val Val Cys Gln Tyr Leu Lys Asn
               305                 310                 315
Pro Asn Val Ser Asp Ser Leu Leu Leu Gln Asn Ile Phe Gln Ala
               320                 325                 330
Leu Asn Val Tyr Tyr Asn Tyr Ser Gly Gln Val Lys Cys Leu Asn
               335                 340                 345
Ile Ser Glu Thr Ala Thr Ser Ser Leu Gly Thr Leu Gly Trp Ser
               350                 355                 360
Tyr Gln Ala Cys Thr Glu Val Val Met Pro Phe Cys Thr Asn Gly
```

```
                365                 370                 375
Val Asp Asp Met Phe Glu Pro His Ser Trp Asn Leu Lys Glu Leu
                380                 385                 390
Ser Asp Asp Cys Phe Gln Gln Trp Gly Val Arg Pro Arg Pro Ser
                395                 400                 405
Trp Ile Thr Thr Met Tyr Gly Gly Lys Asn Ile Ser Ser His Thr
                410                 415                 420
Asn Ile Val Phe Ser Asn Gly Glu Leu Asp Pro Trp Ser Gly Gly
                425                 430                 435
Gly Val Thr Lys Asp Ile Thr Asp Thr Leu Val Ala Val Thr Ile
                440                 445                 450
Ser Glu Gly Ala His His Leu Asp Leu Arg Thr Lys Asn Ala Leu
                455                 460                 465
Asp Pro Met Ser Val Leu Leu Ala Arg Ser Leu Glu Val Arg His
                470                 475                 480
Met Lys Asn Trp Ile Arg Asp Phe Tyr Asp Ser Ala Gly Lys Gln
                485                 490                 495
His
```